THE UNIVERSITY
of EDINBURGH

Generating and characterising Cas12a derived synthetic transcription factors

James William Bryson

This thesis is submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

At the

Institute of Quantitative Biology, Biochemistry and Biotechnology

School of Biological Sciences

University of Edinburgh

2019

THE UNIVERSITY *of* EDINBURGH

Declaration

The following thesis was composed by myself and the work present below is my own unless otherwise stated. This thesis has not been submitted for any other degree or professional qualification.

James William Bryson

Acknowledgements

I'd like to thank Susan Rosser for providing me the opportunity to undertake my PhD in her lab, and for her support throughout the process.

I would like to thank Tessa Moses for guidance, support and being an all round cracking mentor. I would like to thank Dirk Kleinjan for providing early help and protocols regarding the dual Luciferase assay featured in my work. I would like to thank all members of the Rosser lab, in particular, Matt, Pascoe, Jess and Vivek for guidance and support throughout my PhD and thesis writing. Thanks to Trevor Ho for proofreading sections of my thesis and providing helpful advice. Special thanks to Jamie Auxillos, for all the laughs, protocols and academic counsel throughout the years.

I want to thank Elin Ennervald for generously offering her help to establish ChIP after months of troubleshooting.

Finally I'd like to thank my Mum for reviewing my Lay summary despite my comment that I could think of no-one more appropriate.

Abstract

Transcription factors represent one of the primary determinants of gene expression changes and phenotypic modulation during dynamic processes in multi-cellular organisms. This includes processes as diverse and important as; development, differentiation and oncogenesis. Traditional scientific approaches rely on introducing one modification/change at a time and seeing how this impacts a specific outcome. As such, when probing the role from over-expression of a target gene, information can be acquired by over-expressing cDNA derived from a protein encoding gene. However such approaches remove much of the vital biological context provided by the local epigenetic context as well as the role of un-translated regions in modulating the stability, expression and localisation of transcripts/proteins.

Conversely when over-expressing a transcription factor known to target the promoter of a gene of interest, it can become extremely challenging to narrow down the contribution to phenotype caused by the gene of interest being over-expressed, compared to the multiple other gene promoters bound by any one transcription factor. As such the generation of synthetic transcription factors providing predictable, stable binding to a specific locus, with minimal off-target binding provides powerful capabilities for exploration and manipulation of transcriptional networks. To this end, RNA guided homing endonuclease CRISPR systems have been adapted, through the generation of catalytically inactive variants, to serve as easily targetable DNA binding domains. These can be used to recruit transactivation domains to targeted promoters and increase expression for the target gene.

In this work we validate and characterise synthetic transcription factors adapted from Cas12a/Cpf1 derived from three different species. Cas12a, unlike Cas9 possesses the ability

to process its crRNA array through native RNase activity and uses the short crRNAs generated for the targeting of multiple unique loci. This means that designing and constructing genetic constructs for generating guide RNAs becomes cheaper and simpler and furthermore, the reduced size of DNA required for targeting provides benefits when considering packaging size constraints, such as with the AAV virus. We go on to further characterise the capabilities and limitations of these crRNA arrays for a variant derived from *Francisella novicida* - FnCas12a. This variant, unlike the more commonly utilised variants derived from *Acidaminococcus sp BV3L6* (AsCas12a) and *Lachnospiraceae bacterium ND2006* (LbCas12a), requires a shorter characterised PAM sequence 'TTV' for targeting, providing a comparable targeting density to the widely used *Streptococcus pyogenes* derived SpCas9 platforms. Our results open up dFnCas12a as a potential new gold standard for homing endonuclease derived scaffolds for recruiting diverse effector domains, preserving the best qualities of the SpCas9 system, whilst incorporating added strengths alongside the innate ability to operate orthogonally to existing SpCas9 systems.

Lay Summary

Almost all of the cells in our body share the same DNA sequence with the same genes and yet they come in various shapes, sizes and functions, helping to make your heart beat, or filter toxins out of your blood. Your genes each produce unique proteins that help define what a cell looks like, how it functions and interacts with its neighbours. It is the differences in expression of these genes and the proteins they produce that explains the extraordinary diversity of cells we see.

My work focusses on trying to create more effective tools to cause over-expression of multiple chosen genes. This is very useful for scientists trying to better understand how these genes work together to cause changes in the cell. Alternatively for cell therapies, where we need to deliver a specific cell type to help repair an organ, this tool can allow us to take cells from our skin and over-express several genes to change the cell from one type to another.

Finally with genetic conditions such as sickle cell anaemia, the version of the oxygen carrying protein 'haemoglobin' produced by adults is defective, but this can be fixed by increasing expression of a different version that is only well expressed until we are 4 months old. This is a relatively simple solution to a well known genetic condition, however many genetic conditions are caused by problems associated with multiple genes. The tool I have developed in the following work is especially well suited for specifically interacting with multiple genes and causing their over-expression. We have shown the upregulation of multiple genes with targeting achieved from a single DNA construct. Furthermore we have provided information on the capabilities and limitations of targeting, which can help aid biological researchers and potentially clinical researchers.

It is one thing to better understand the complicated causes of genetic conditions, it is quite another to use this information to try and fix debilitating conditions. The tool I have been working on should be able to help address both of these. First by allowing us to more easily change expression of multiple chosen genes for cells in a dish and gaining insight by seeing how this impacts the cells and whether we see desirable changes. Secondly after rigorous testing this tool could be used alongside the information we have learned to try and tackle what are currently incurable, debilitating genetic conditions.

Table of Contents

List of figures

List of tables

Chapter 1 Introduction

1.1 Synthetic transcription factors

Transcription factors are found in both prokaryotes and eukaryotes and help coordinate transcriptional regulation, by increasing the transcription rate of a physically proximal gene through the recruitment of RNA polymerase to an associated promoter. Transcription factors typically target multiple loci in a context specific fashion (Wang, 2005) (Teif and Rippe, 2010), with binding to a locus being driven by chromosomal accessibility and the presence of the 'consensus sequence' within DNA. Unlike those employed for synthetic transcription factors, the DNA binding domains of natural transcription factors typically have highly specific binding to only 5-8 nucleotides within a short stretch of DNA, alongside less specific binding to surrounding nucleotides. While this provides flexibility within an evolutionary context, it creates challenges for researchers wishing to induce specific effects at targeted loci in a controlled fashion, creating demand for more precise and predictable binding domains. As such, researchers have explored synthetic transcription factors as a means of specifically transactivating target genes.

One of the first synthetic transcription factors to be successfully developed utilised zinc-fingers (Liu et al., 1997) (Figure 1-1A), with the design centring around zinc finger motifs naturally found in many transcription factors. Initially zinc finger approaches seemed very promising, with early work presuming a simple 'mix and match' of zinc fingers individually designed with specificity for each of the DNA codons, thereby requiring only 64 independent zinc-fingers to enable specific coverage of any DNA sequence (Pavletich and Pabo, 1991). There were, however, multiple caveats. Not only was the construction very

laborious and only carried out by specialised labs, but perhaps the biggest limitation was the cross-recognition of adjacent zinc fingers. This means that the binding capability of individual zinc finger motifs are
dependent on the adjacent zinc finger motifs, complicating the design and necessitating experimental screening of binding. As such, this technique has rapidly been outpaced by more recent techniques mentioned below.

In the following years a new approach utilising 'transcription activator-like effectors' (TALEs) was discovered. TALEs are a virulence factor produced by *Xanthomonas* (a proteobacteria that causes disease in plants), which mimics eukaryotic transcription factors to reprogram host cells (Zhang et al., 2000). Their targeting capabilities were first characterised in 2009, (Boch et al., 2009), and the same team also showed that TALEs could be used to construct artificial DNA binding proteins capable of targeting any desired sequence. The TALEs DNA binding domain is often composed of tandem repeats of 34 amino acid modules, each possessing variability at positions 12 and 13. It is these 'repeat variable di-residues' (RVDs) that target a specific base of the DNA (Figure 1-1B) (Deng et al., 2012) (Morbitzer et al., 2010)(Streubel et al., 2012). These RVD repeats can be 5.5 to 34.5 repeats long with the final 0.5 repeat contributing to the correct binding of the last nucleotide target (Boch et al., 2009). The TALE system offers unique advantages over the Zinc finger approaches, as each module is able to bind independently of the surrounding modules, meaning that TALEs could be predictably designed.

The following year, Morbitzer *et al.* demonstrated that TALEs could be used to generate synthetic transcription factors (Morbitzer et al., 2010). In particular, they showed that by changing the RVD's within the TALE-Avrbs3 fusion protein, they were able to target the protein to induce robust up-regulation of the Bs4S gene, measured using semiquantitative

RT-PCR. Since then, TALE based synthetic transcription factors have been employed for a number of uses including the treatment of genetic conditions (Chapdelaine et al., 2013), reprogramming mouse fibroblasts to pluripotency (Gao et al., 2013) and regulating synthetic gene circuits (Garg et al., 2012).

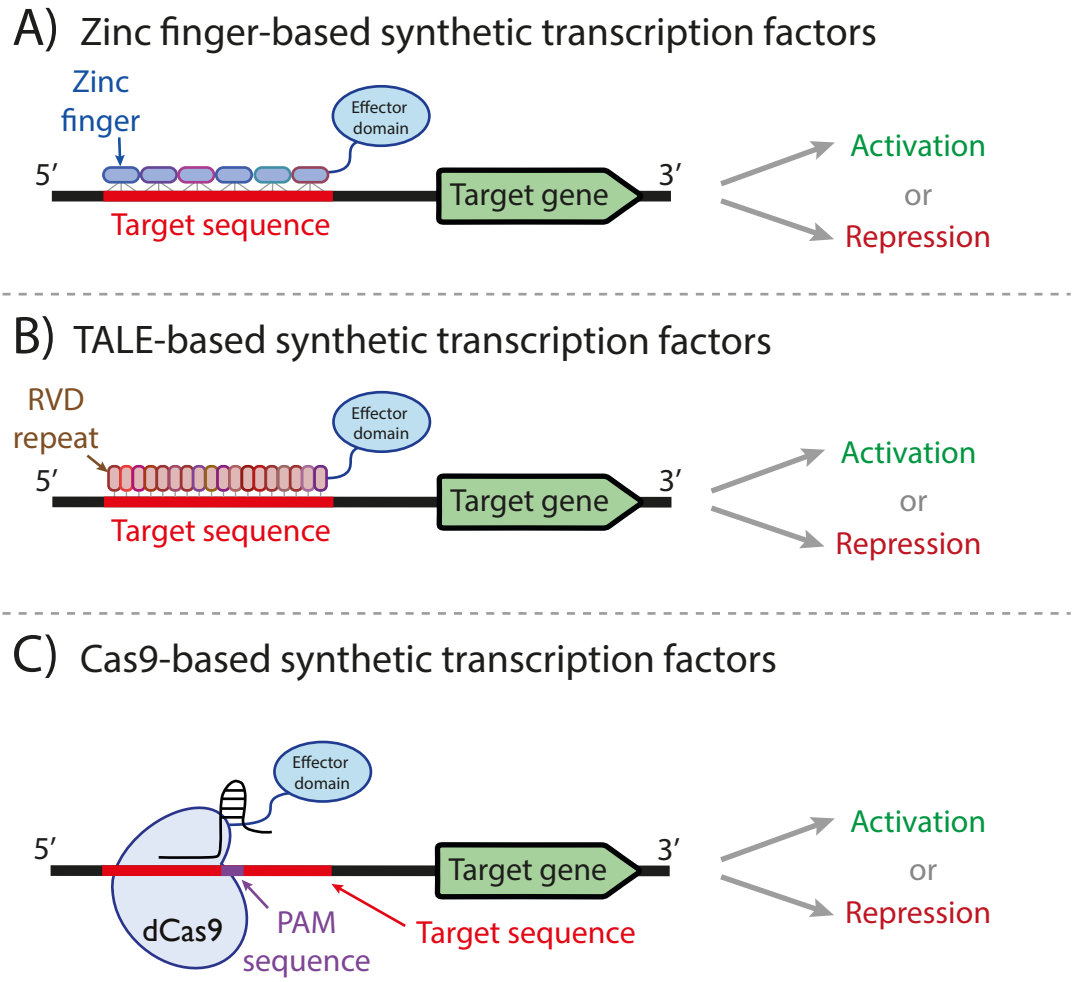

Figure 1-1 - Evolution of synthetic transcription factors

Diagrammatic representation of the zinc finger, TALE and Cas9 derived synthetic transcription factors/transcriptional repressors. For zinc finger-based synthetic transcription factors, each zinc finger is able to recognise 3 nucleotides, dependent upon the adjacent zinc fingers. For TALE-based synthetic transcription factors, each RVD repeat can bind and recognise a single nucleotide independent of adjacent repeat variable di-residues (RVD) repeats. For Cas9-based synthetic transcription factors, the dCas9-effector protein can be guided to a locus, as long as an adjacent PAM sequence is present and the spacer sequence within the gRNA can hybridise the target sequence

1.2 CRISPR/Cas9

One of the most transformative innovations for DNA binding modules and synthetic transcription factors has been the discovery of the clustered regularly interspaced short palindromic repeats (CRISPR) Cas9 system, and its subsequent adaptation for use as a synthetic transcription factor. CRISPR/Cas9 evolved as a form of adaptive immunity, which can process and integrate foreign DNA into the host genome and use the newly produced transcripts as a targeting mechanism to specifically cleave the foreign DNA (Figure 1-2).

The first step of this process, known as adaptation, relies upon the processing of foreign DNA by Cas1 and Cas2 to produce a short fragment of DNA that is subsequently inserted into the CRISPR array (Yosef et al., 2012) (Figure 1-2). This (often viral) 'protospacer' DNA must possess a protospacer adjacent motif (PAM), adjacent to the protospacer, that can then be recognised by the Cas9 protein in down-stream cleavage reactions. However this PAM sequence is not incorporated into the crRNA array to prevent self-targeting (Horvath et al., 2008). In the second stage – CRISPR RNA (crRNA) processing, the precursor crRNA array is transcribed and processed into mature crRNA. First trans-activating crRNA (tracrRNA) hybridises with the direct repeats within the crRNA array, Cas9 and RNaseIII can subsequently process the crRNA arrays into independent guide RNAs (gRNA). In the final interference step, the gRNA combine with the Cas9 protein, to enable targeting of foreign DNA. When the mature Cas9/gRNA complex interacts with the appropriate PAM sequence ('NGG' for Spcas9) the DNA is melted and if the adjacent protospacer sequence is complementary to the spacer sequence within the crRNA then the Cas9 binds and induces a blunt ended double strand cleavage of the target DNA (Figure 1-2). CRISPR/Cas9 is a Class 2 system, as, unlike Class 1 systems, only a single multidomain protein (Cas9) is required for interference.

Whilst the direct repeats, that are characteristic of CRISPR were initially identified in 1987 (Ishino et al., 1987), it was only in 2007 that these repeats were characterised to play a role in bacterial defence against viruses (Barrangou et al., 2007). Subsequent work showed that Cas9 was the only *cas* gene essential for CRISPR-encoded interference when expressed in *Escherichia coli (E. coli)* (Sapranauskas et al., 2011). In 2012 the promise of the CRISPR system and in particular Cas9 was highlighted by Jinek and colleagues (Jinek et al., 2012). The team showed that Cas9 could be targeted to specific DNA sequences using only a single guide RNA, composed of a crRNA-tracrRNA fusion, as opposed to the two separate crRNA and tracrRNA molecules normally required for successful targeting in bacteria. Initially, this innovation represented the capability to easily construct gRNAs to target the wild-type Cas9 endonuclease to desired chromosomal loci, inducing double strand cleavage. This enabled either the formation of indel mutations through inefficient non-homologous end joining (NHEJ) in human cells (Mali et al., 2013), or the targeted insertion of genetic material using a homologous repair template through homology directed repair (HDR) in mouse cells (Yang et al., 2013).

Jinek et al. also showed that introducing single amino acid substitutions in the two Cas9 nuclease domains HNH and RuvC lead to their inactivation (Jinek et al., 2012). This catalytically inactive “dead Cas9” (dCas9) maintained the capability to bind DNA but no longer induced cleavage (Figure 1-3A) and could serve as a simple RNA-guided DNA binding domain which could be used as a scaffold for synthetic transcription factors, repressors or chromatin editors.

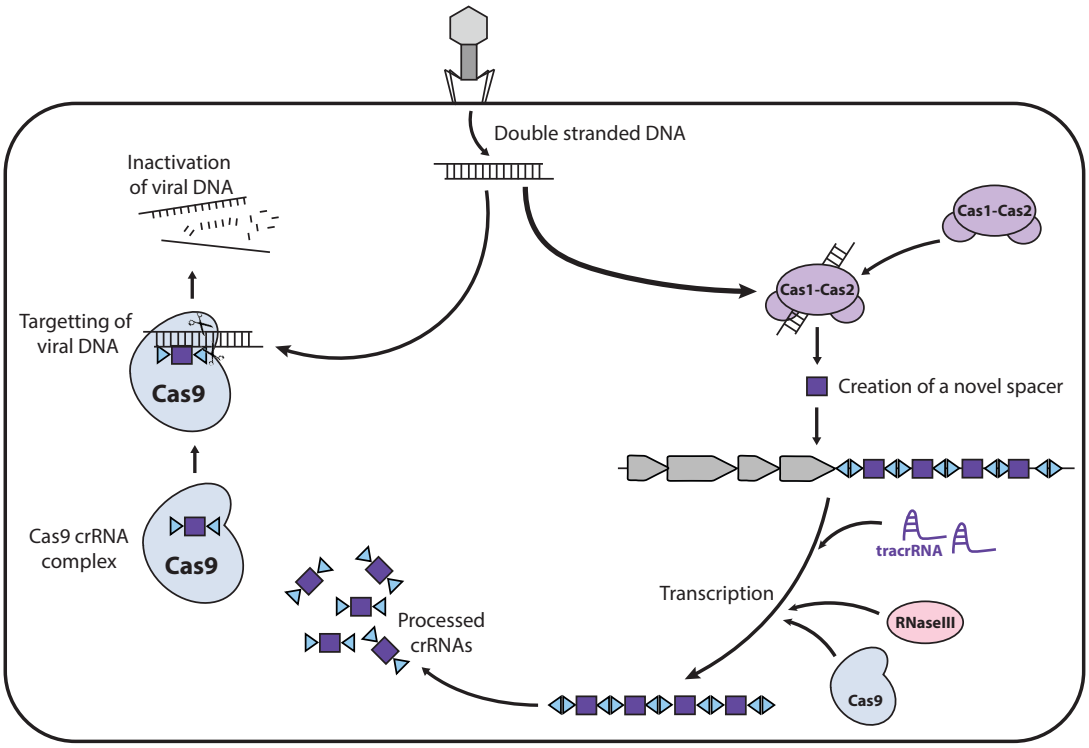

Figure 1-2. - Representation of CRISPR/Cas9 adaptive immunity

In stage 1, invading DNA is processed by the Cas1-Cas2 complex. A processed spacer sequence can then be incorporated between two direct repeats within a crRNA array. In stage 2 the crRNA array is transcribed and processed, with tracrRNA hybridising to the direct repeats, enabling processing by RNaseIII and Cas9 to generate individual gRNAs (with one spacer sequence containing crRNA per gRNA). In the final stage, the gRNA can complex with Cas9, the Cas9 protein can recognise and melt a protospacer-adjacent motif (PAM) sequence and if the spacer sequence is able to hybridise with the melted DNA, Cas9 induces a double strand break in the target DNA.

1.3 CRISPR/Cas9 transcriptional repressors

One of the early observations for the dCas9 protein was its ability to induce transcriptional repression in bacterial and mammalian cells by providing a steric block to the processive RNA polymerase (Gilbert et al., 2013)(Qi et al., 2013). However, while this system allowed approximately 1000 fold silencing in E.coli, only approximately 2 fold repression was achieved in mammalian cells. As such researchers generated fusion proteins to act as transcriptional repressors. Gilbert and colleagues tested three previously characterised repressor domains using the dCas9 scaffold; the Krüppel associated box (KRAB) (Margolin et al., 1994), the CS domain from HP1α (Hathaway et al., 2012) and the WPRW domain from Hes1 (Fisher et al., 1996). The team saw that the KRAB-dCas9 fusion protein was the most effective at repression, inducing 15-fold repression when targeted to GFP expressed from a constitutive promoter in HEK293 cells that had been delivered by lentivirus 14 days prior (Gilbert et al., 2013). The capabilities of transcriptional repressors such as dCas9-KRAB make for very powerful tools when combined with synthetic transcription factors, by enabling simultaneous transactivation and repression of unique genes, enabling interaction with gene networks that frequently incorporate complicated feedback loops, relying on repression as well as activation.

1.4 CRISPR/Cas9 synthetic transcription factors

Within a year multiple groups showed the utility of this scaffold, attaching and testing activator domains such as the VP16 tetramer - VP64 (Maeder et al., 2013a) and p65 (Gilbert et al., 2013) (Figure 1-3B). Gilbert and colleagues separately fused VP64 and p65 activation domains to dCas9 proteins, which were subsequently targeted to the upstream activation sequence of the Gal4UAS-GFP reporter. When measuring GFP levels two days post transfection, the team saw a 25-fold up regulation of GFP for the dCas9-VP64 fusion

protein relative to a negative control with no targeting gRNA. They also observed a 12-fold up regulation of GFP for the dCas9-p65 fusion protein relative to a negative control with no targeting gRNA. They utilised a similar GFP reporter strategy for analysing repression and tested the effect of the dCas9-KRAB fusion protein when targeted to multiple different loci within an SV40 promoter driving expression of a GFP reporter gene. They observed up to a 15-fold repression of GFP expression when targeting the transcription start site.

An important observation Maeder and colleagues noticed when inducing activation of the *VEGFA* and *NTF3* genes, was improved transactivation when targeting multiple rather than single gRNAs to the respective promoter. They observed that the up-regulation of expression induced using multiple gRNAs was significantly higher than predicted from the additive effects of all the single gRNAs acting independently. This was in agreement with what had been learnt from earlier TALE-based approaches (Maeder et al., 2013b). It is important to note that, whilst more effective, the use of multiple gRNAs to target the same locus would be less desirable for projects where many loci are targeted, due to the increased number of gRNAs and associated challenges for design and delivery. In part to address this problem, subsequent work has focussed on generating synthetic transcription factors possessing improved activity.

1.5 More powerful effectors

Work focussing on generating stronger and more reliable transcriptional activators has led to a rapid expansion of verified dCas9 transactivators, whether they rely on previously characterised domains (Maeder et al., 2013a) (Gilbert et al., 2013) or trial alternative variations such as the VPR domain (Chavez et al., 2015). The VPR domain is a three part effector comprised of the transactivator domains; VP64, p65 and Rta connected by flexible glycine-serine linkers, and fused to dCas9 by such a linker (Figure 1-3B). This effector

allowed greater increases in expression of a range of genes when compared to VP64 alone, and as of writing remains the best single component activator directly coupled to dCas9. Chavez et al. observed a 22- to 320-fold improved activation of targeted genes relative to the dCas9-VP64 fusion protein when relative RNA expression was analysed. Furthermore, the team demonstrated that the approach was suitable and effective for multiplexing, with the genes *MIAT*, *NEUROD1*, *ASCL1* and *RHOXF2* showing simultaneously significantly higher gene expression for all tested genes relative to a dCas9-VP64 control. One caveat of the study was the group's reliance on 3-4 gRNAs per promoter to induce up-regulation. Further testing using only one gRNA per targeted promoter could help demonstrate the broader utility or limitations of this effector.

Other approaches have also been devised in an effort to maximise expression, with more complex two component activators such as the SunTag (Tanenbaum et al., 2014) and SAM systems (Konermann et al., 2015) being developed. The SunTag approach relies on fusing a tandem array of peptides to the dCas9 protein (Figure 1-3C). These peptides can be used to recruit single-chain variable fragments, engineered antibody fragments, which are fused to the VP64 transactivator. Using this approach, one dCas9 protein can recruit up to 24 copies of VP64, reliably increasing activation far beyond single copy variants. This in turn translated to not only a significant up-regulation of the chemokine receptor *CXCR4* gene, but they observed that this up-regulation resulted in a corresponding change in migration of the K562 cells used in the study. The cells showed a less than 2-fold increase in migration when the gene expression was up-regulated using dCas9-VP64 compared to over 15-fold increase in cell migration when the gene expression was up-regulated using dCas9-SunTag-VP64.

A further approach, the synergistic activation mediator (SAM) system (Figure 1-3C), also relies upon recruitment of multiple copies of the activators p65 and HSF1, whilst also tethering a VP64 domain fused directly to the dCas9. This system however utilises RNA-aptamer based recruitment, whereby the gRNA is designed to incorporate two MS2 hairpin loops within the tetraloop and stem-loop 2 (Konermann et al., 2015). Konermann and colleagues went on to show for 10 out of the 12 genes tested the SAM system showed stronger activation than pools of 8 single gRNAs used in conjunction with dCas9-VP64. The SAM system also enables gRNA dependent orthogonality, as changing the aptamer incorporated into the gRNAs enables recruitment of alternative aptamer binding proteins, which in turn can be tethered to antagonistic effectors, as later shown by Truong et al. where they showed simultaneous transactivation and repression (Truong et al., 2019).

When comparing the relative strengths and weaknesses of the previous approaches, it is important to highlight that whilst two component systems such as SunTag and SAM can offer increased recruitment of transactivation domains or intrinsic orthogonality, respectively, the necessity for co-delivery of secondary effectors (either linked to antibody fragments or aptamer binding proteins) complicates the design and delivery of necessary constructs.

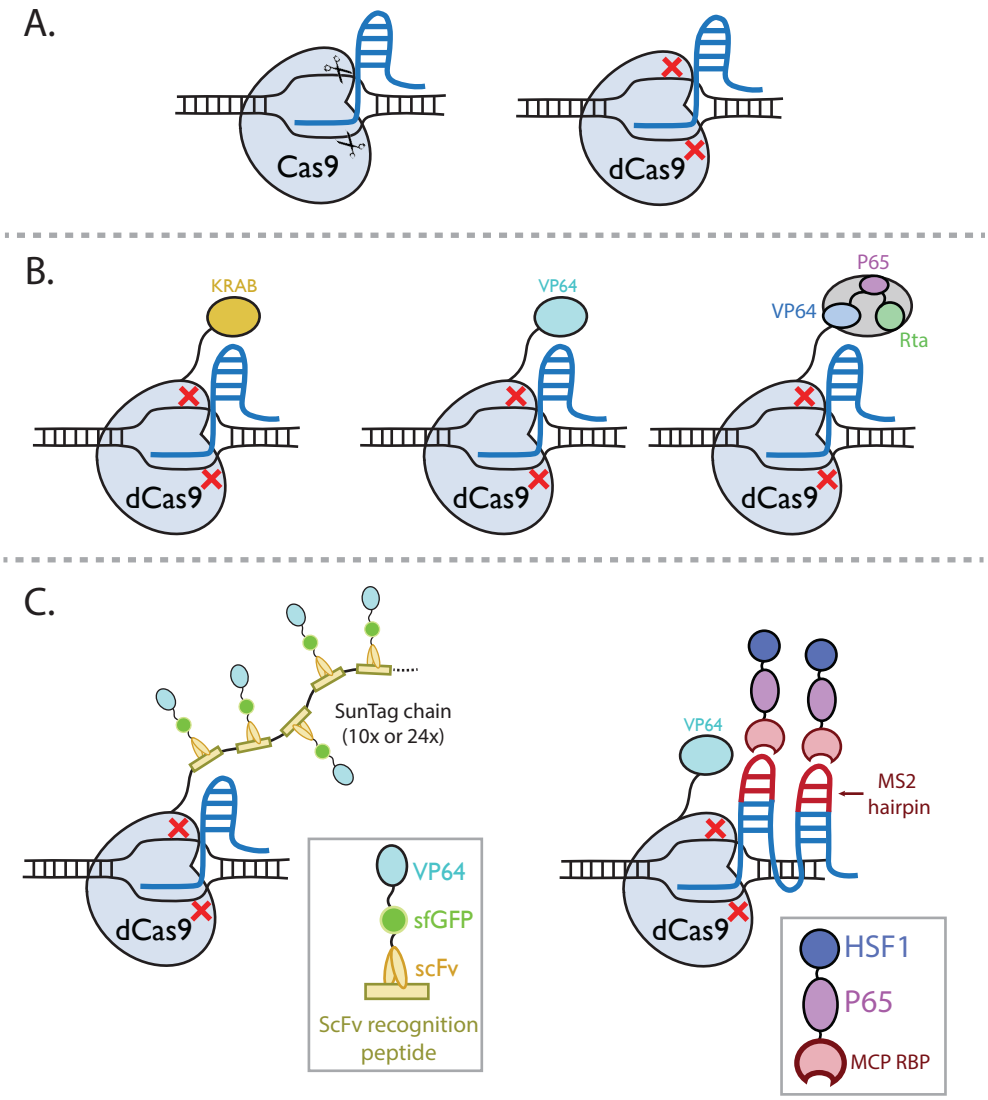

Figure 1-3 - Classes of Cas9-based synthetic transcription factors

A) Wild type Cas9 can be targeted to a locus by a gRNA to induce a double strand break. Through two amino acid substitutions in the HNH and RuvC domains, cleavage can be abolished creating a RNA guided DNA binding domain (dCas9). B) Single domain or multi-domain effectors can be fused to dCas9, to enable targeted transactivation or repression of adjacent genes using a single fusion protein. C) Multi-component approaches rely on the recruitment of secondary proteins to either recognition peptides (SunTag system) or to aptamers incorporated within the gRNA (SAM system).

1.6 Multiplexing

With the exception of the SAM system, current approaches rely on effectors fused or recruited directly to the dCas9 protein. This means that when utilising the same dCas9 variant, only one effector, such as an activator or a repressor, can be utilised within a cell as all targeting gRNAs will interact with the dCas9 irrespective of the fused effector domain. This creates some important challenges for areas such as cellular reprogramming, where complex gene networks rely on negative as well as positive feedback loops (which might require repressors and activators, respectively) and the ability to inhibit gene expression may prove just as important as the ability to activate master transcription factors. As such, the need has arisen for orthogonal systems, which enable the simultaneous targeting of a synthetic transcription factor to a subset of loci and the targeting of a transcriptional repressor to a different subset of loci, without cross-talk occurring. This need has been met by a number of strategies.

One such strategy relies upon the intrinsic orthogonality of the TALEs, which function in a gRNA independent way. However, utilising TALEs leads to the previously described limitations, where individual proteins must be designed and generated for each locus to be targeted. An alternative strategy that has been developed relies upon the subtle differences between the Cas9 variants found in nature and the orthogonality of their respective gRNAs. For example, Esvelt and colleagues showed that three Cas9 variants, from *Streptococcus pyogenes* (dSpCas9), *Streptococcus thermophilus* (dStCas9) and *Neisseria meningitidis* (dNmCas9), showed high orthogonality relative to one another (Esvelt et al., 2013). In particular when

Further work by Gao and collegues showed that simultaneous activation and repression of two different reporter genes could be achieved when employing a dSpCas9-KRAB repressor

alongside an *Staphylococcus aureus* dSaCas9-VPR synthetic transcription factor (Gao et al., 2016). However, there is a reason why *S. pyogenes* derived Cas9 is most consistently used and favoured for targeting, as it possesses a very short PAM sequence 'NGG', allowing targeting on average every 8bp of a genome when targeting either strand. This contrasts with the PAM sequences seen for *Staphylococcus aureus* (SaCas9), *Neisseria meningitidis* (NmCas9) and *Streptococcus thermophilus* (StCas9) with PAM sequences; 'NNGRRT', 'NNNNGATT' and 'NNAGAAW' enabling targeting every 32, 128 and 256 bp respectively. Whilst such limitations may be acceptable for some applications, they can pose a problem for targeting multiple promoter regions with synthetic transcription factors, where positioning can be essential, with evidence showing an optimal window of 50 to 400bp upstream of transcription start sites (Gilbert et al., 2014)

Furthermore, targeting limitations can provide a particular challenge when targeting promoters with several gRNAs, which has been demonstrated to be more effective in activation than single gRNA approaches (Maeder et al., 2013a) (Perez-Pinera et al., 2013).

1.7 Chromatin editors

Alongside the transcriptome, researchers have also explored modification of the epigenome – the semi-heritable chemical changes to genomic DNA and associated histones. Our understanding of the roles of specific histone and DNA modifications is only increasing with time (Snowden et al., 2002) (Costello et al., 2013) and large scale projects are continuing to characterise epigenomes from multiple cell types (Roadmap Epigenomics Consortium et al., 2015). Global epigenetic modifications through the delivery of chemicals (Halby et al., 2012) or over-expression/miRNA knockdown of native epigenetic modifiers (Luco et al., 2010) have enabled correlations between epigenetic modifications and a number of phenotypes to be established. However, addressing narrower questions of

causation and whether the presence of a mark at a specific locus is causing a specific phenotype requires more specific, targeted approaches.

As a result, over the last few years much work has been put into the creation of specific chromatin editors as a more targeted and potentially informative approach relative to the global epigenetic changes induced by chemical approaches (Cole, 2008). Similar to the problems faced with the construction of synthetic transcription factors, one of the key issues has been identifying and testing appropriate effector domains, in this case for inducing targeted epigenetic modifications.

One of the earliest chromatin editors employed as a DNA binding fusion protein was originally tested and characterised a transcriptional repressor protein (Deuschle et al., 1995). The KRAB domain, was subsequently shown to mediate its effects through heterochromatin formation, through the recruitment of proteins capable of laying a trimethyl mark on histone 3, lysine 9 (H3K9me3) (Groner et al., 2010; Ryan et al., 1999; Schultz et al., 2002; Sripathy et al., 2006) (Figure 1-4A). As such the KRAB domain also possesses uses beyond direct transcriptional repression, for example a recent paper highlighted the use of KRAB to repress distal enhancers (Thakore et al., 2015). Thakore and colleagues showed that targeting of dCas9-KRAB to the HS2 distal enhancer down-regulated the expression of the global genes *HBE1* and *HBG1*/*2*, after quantifying relative expression with qRT-PCR. Alternative repressors have also been developed, in part due to the fact the KRAB domain can induce repression when targeted to distal or proximal elements associated with a gene (Gao et al., 2014; Gilbert et al., 2014, 2013; Kearns et al., 2015). LSD1 has been characterised to induce demethylation of mono and di-methylated lysine 4 and 9 on the H3 histone (Metzger et al., 2005; Shi et al., 2004). By targeting LSD1 using a TALE and dCas9 fusions, Kearns and colleagues were able to demonstrate robust

repression of *TBX3* when targeted to the distal enhancer Enh1 but not when targeting to the proximal promoter (Kearns et al., 2015).

Through the adaptation of the histone acetyltransferase p300 (Ogryzko et al., 1996), researches have also demonstrated the capacity to induce an open chromatin state at targeted loci, with dCas9 based recruitment of the p300 catalytic core (Hilton et al., 2015) (Figure 1-4B). They also went on to show that they could induce transactivation through the targeting of a distal enhancers for *MYOD* and *OCT4* as well as the β-globin locus.

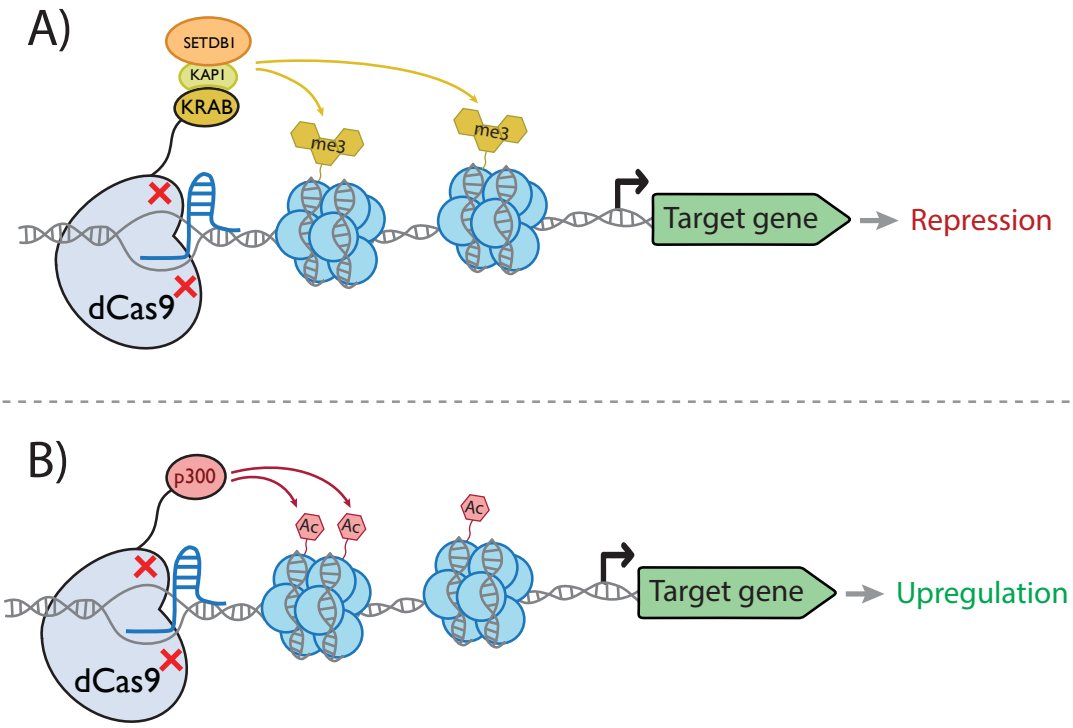

Figure 1-4 - Synthetic chromatin editors

A) Diagrammatic representation of the action of dCas9-KRAB to enable laying of the formation of heterochromatin, leading to gene repression. This is achieved through the recruitment of secondary proteins KAP1 and SETDB1, which lay the histone 3 lysine 9 trimethyl mark on proximal nucleosomes. B) Diagrammatic representation of the action of dCas9-p300. As a histone acetyl-transferase, p300 directly acetylates proximal histones, leading to increased turnover of histones and a transition towards a euchromatic state, enabling increased gene expression.

The previously described chromatin editors modified the histones that DNA associates with. However, researchers have also tethered domains onto dCas9, that directly modify the methylation of DNA, associated with transposon silencing (Goll and Bestor, 2005) and repression of promoters (Suzuki et al., 2007). Tet1 DNA demethylase has been tethered to dCas9 and targeted to the *FMR1* gene to demethylate the CGG repeats found in the 5'UTR, responsible for Fragile X Syndrome. Liu and colleagues showed that they could observe sustained DNA demethylation and increased gene expression over a two week period, with rescue of the hyperactive electrophysiological phenotype observed for edited patient derived neurons (Liu et al., 2018). In contrast, by fusing the DNA methyltransferase Dnmt3a to dCas9 and targeting it to a promoter driving GFP expression, they observed both an increase in methylation of the promoter region, alongside a corresponding increase in GFP negative cells (Liu et al., 2016).

1.8 Cas12a

It is important to remember when considering CRISPR approaches that Cas9 represents only one of several functional gRNA driven endonucleases. In fact with the discovery and further characterisation of Cas12a (formerly Cpf1) (Zetsche et al., 2015a), it appears that alternative CRISPR systems may offer unique advantages that can enable them to complement and for some applications supersede the existing Cas9 derived toolkits.

Similarly to Cas9, Cas12a is a homing endonuclease, relying upon a crRNA sequence to confer target specificity and like Cas9, Cas12a is a class 2 CRISPR system, with interference activity being achieved by a single multi domain protein (Figure 1-5A). However, initial characterisation of Cas12a by Zetsche and collegues showed Cas12a possessed a number of unique properties (Zetsche et al., 2015b). They observed that Cas12a does not to require a tracrRNA for targeting. They also showed that in contrast to Cas9 DNA cleavage by Cas12a

produces a staggered break with a 5' overhang and is initiated by a RuvC-like domain (Figure 1-5A). Unlike for Cas9, which requires two independent amino acid substitutions to abolish DNA cleavage, a single amino acid substitution in the RuvC domain (E993A) can inhibit cleavage for both strands. For the 16 Cas12a variants they screened, they were able to identify clear PAM sequences for 8 and went on the show cleavage in mammalian cells for two of these variants: *Acidaminococcus sp* (AsCas12a) and *Lachnospiraceae bacterium* (LbCas12a). These were originally characterised to have PAM sequences of TTTN, however subsequent publications have refined this, demonstrating that the two variants optimally target a TTTV PAM sequence (Kim et al., 2017).

After initial characterisation by Zetsche and collegues, most subsequent application of Cas12a have focussed on the As and Lb variants, as the original work suggested only these variants could successfully cleave DNA in mammalian cells (Zetsche et al., 2015a). AsCas12a has been applied for editing humanised liver cells in mice (Tsukamoto et al., 2018) and mice neurons (Zetsche et al., 2017) through *in vivo* delivery with adeno-associated virus (AAV). LbCas12a has been applied for the generation of an APOE knockout rat, serving as a model for atherosclerosis (Lee et al., 2019) as well as for the correction of dystrophin mutations responsible for muscular dystrophy in both human cardiomyocytes and mice (Zhang et al., 2017). Of interest, despite initial analysis by Zetsche *et al.* showing activity in human cells for only the As and Lb variants, work by Tu *et al.* showed that the *Francisella tularensis subsp. novicida U112* derived Cas12a (FnCas12a) also showed robust cleavage activity when targeting three different genomic loci across three different genes (Tu et al., 2017). This was also extended to *Moraxella bovoculi 237 derived* Cas12a (MbCas12a), with work by Tóth *et al.* which showed that non-homologous recombination and homology directed recombination could be achieved for the Fn and Mb variants alongside the As and Lb (Tóth et al., 2018).

A number of studies have been conducted to enhance the capabilities of the wild-type variants of Cas12a and to improve cleavage activity and targeting capabilities for Cas12a. Park *et al.* showed that by extending the 5' of the crRNAs, they were able to observe increased gene editing when targeting reporter constructs in HEK293 cells (Park et al., 2018). Gao *et al.* were subsequently able to generate two variants of AsCas12a ('RR' and 'RVR') that could target TYCV and TATV PAM sequences, as opposed to the TTTV PAM sequence for the original AsCas12a (Gao et al., 2017). Finally, drug inducible cleavage has been enabled for AsCas12a through the insertion of the oestrogen receptor domain into a flexible loop, with cleavage only being observed when tamoxifen was present (Dominguez-Monedero and Davies, 2018).

A number of tools have now been developed to facilitate crRNA selection and increase the likelihood of successful genome editing. These include work by Park and colleagues who developed software to enable rapid design of crRNAs for genome-wide screening experiments, using As and Lb variants (Park et al., 2018). Kim *et al.* went on to utilise a large dataset of AsCas12a cleavage activity and chromatin accessibility data to train a deep-learning algorithm to facilitate the selection of crRNAs with maximised activity (Kim et al., 2018).

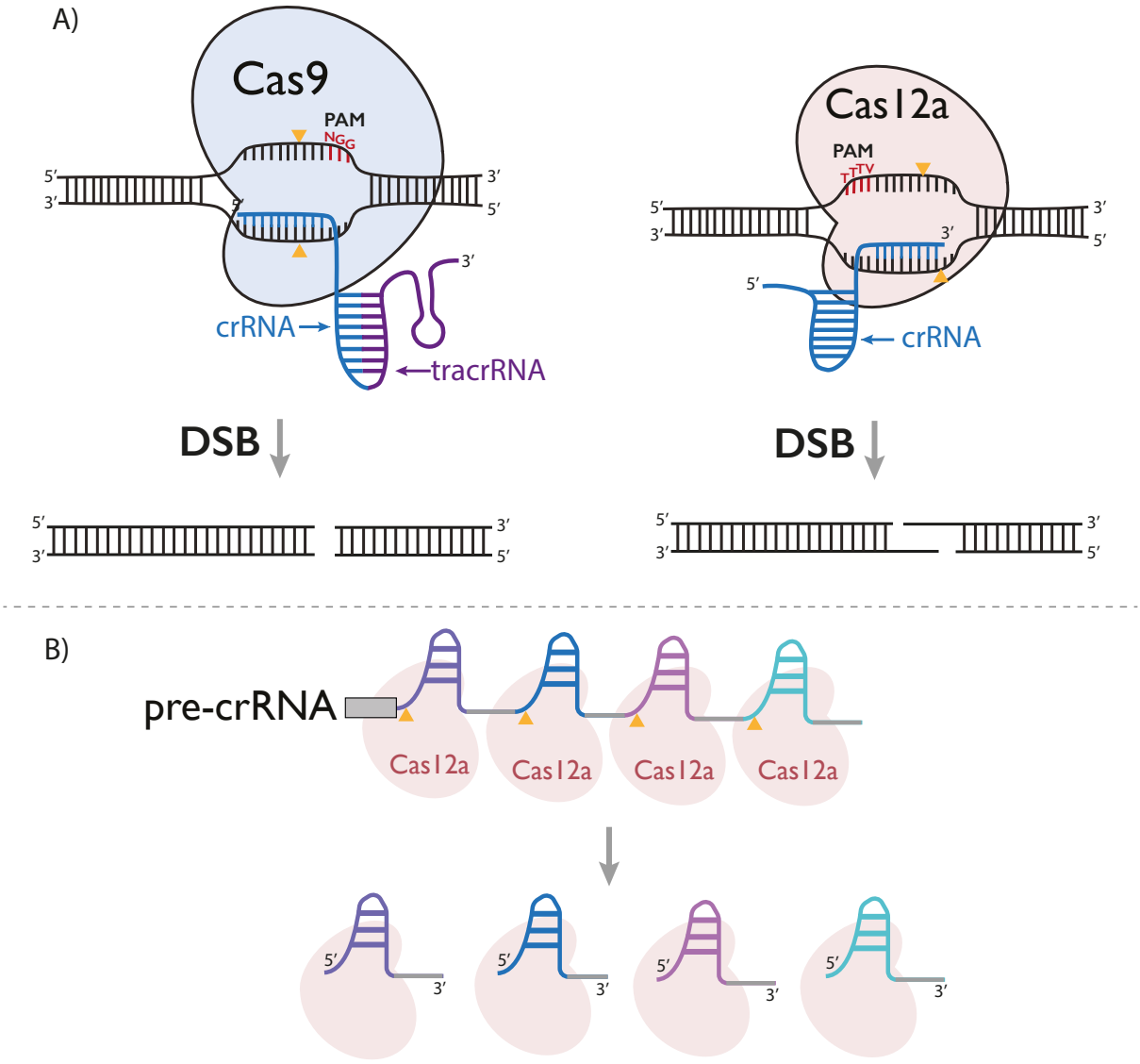

Figure 1-5 - Comparison of Cas9 and Cas12a

A) Diagrammatic representation of some of the key differences between Cas9 and Cas12a. Cas9 recognises a G rich PAM sequence to the 3' of the protospacer sequence and melts the DNA to enable hybridisation by the spacer sequence of the crRNA (fused to a tracrRNA) before inducing a blunt double-strand break. Cas12a recognises a T rich PAM sequence to the 5' of the protospacer sequence and melts the DNA to enable hybridisation of the spacer sequence of the crRNA, before inducing a staggered 4/5bp double strand break. B) Cas12a unlike Cas9 possesses RNase activity that enables it to process it's own crRNA array, enabling a single crRNA array transcript expressed alongside Cas12a to generate single crRNAs for subsequent targeting.

1.9 *crRNA arrays*

Initial work by Zetsche *et al.* also showed that crRNA arrays, composed of multiple tandem targeting crRNAs, could be processed when only Cas12a and the crRNA array were expressed, without further Cas proteins or tracrRNA (Zetsche et al., 2015a) (Figure 1-5B). Subsequent work by Fonfara *et al.* went on to show that FnCas12a possesses RNase activity and is sufficient to process crRNA array *in vitro* (Fonfara et al., 2016). This ability to process a crRNA array provides a greatly simplified strategy for the delivery of multiple gRNAs, when compared to Cas9. For Cas9 systems gRNAs must either be expressed from multiple transcriptional units or using more complicated strategies relying upon co-delivery of an endoribonuclease (Nissim et al., 2014), or the flanking of the gRNAs with ribozymes/tRNA (Nissim et al., 2014)(Zhang et al., 2017) genes to enable appropriate processing.

A number of groups have showcased the strength of using arrays for targeting, with applications for multiplexed gene editing (Zetsche et al., 2017) as well as synergistic and multiplexed transactivation observed in human cells (Tak et al., 2017)(Zhang et al., 2018) alongside multiplexed repression (Campa et al., 2019).

1.10 *Cas12a repression*

Alongside work with wild-type Cas12a, researchers have also explored the adaptation of CRISPR/Cas12a for repression of gene expression. Utilising the E993A mutation originally characterised by Zetsche *et al.* to abolish cleavage activity, Zhang *et al.* showed that gene repression could be achieved by employing dAsCas12a in *E. coli* (Zhang et al., 2017), similar to what had previously been shown in *E. coli* using dCas9 (Gilbert et al., 2013). However, similarly to observations for dCas9, minimal repression was observed when dAsCas12a was

targeted, with a 2-fold repression being achieved only when three crRNAs were simultaneously targeted to a CMV promoter controlling the expression of GFP (Liu et al., 2017). The same publication showed that by tethering the KRAB domain to the C-terminus of dAsCas12a, higher than two-fold repression could be achieved with single crRNAs, which became further pronounced when multiple unique crRNAs were co-delivered to the HEK293T cells.

1.11 Cas12a synthetic transcription factors

Researchers also demonstrated the capacity of dCas12a scaffolds to enable recruitment of transactivation domains to genomic loci to serve as synthetic transcription factors. Initial work by Tak and colleagues demonstrated that by tethering either the p65 or VPR domain to dLbCas12a, they could observe significant transactivation for targeted genes in HEK293 cells, however this transactivation was significantly higher when the VPR domain was employed (Tak et al., 2017). In the same study they also demonstrated the use of the chemically inducible dimerisation domains DmrA and DmrC (Rivera et al., 2012), which recruited the two individual effector domains to dLbCas12a when the A/C heterodimeriser compound was present. Finally they also went on to employ a series of 3-crRNA arrays. When comparing the transactivation observed for an array possessing three crRNAs targeting dLbCas12a-VPR to a single promoter, compared to the transactivation enabled by each individual crRNA, they were able to see not only higher, but synergistic transactivations for two of the three genes tested. This meant that the crRNA arrays were able to induce a greater transactivation than the predicted additive effect from combining each of the single crRNAs. Subsequent work has confirmed the efficacy of dLbCas12a-VPR. Zhang and colleagues not only showed that dLbCas12a-VPR was able to more robustly transactivate compared to dLbCas12a across a number of targets (Zhang et al., 2018).

The capacity for dAscCas12a-VPR to also induce robust transactivation in mammalian cells was demonstrated by Liu and colleagues, who also observed increased activity when a pool of three crRNAs were targeted compared to the individual crRNA (Liu et al., 2017). They went on to show that through the incorporation of either a tetracycline or theophylline aptamer in the 3' end of the crRNA, they were able to generate druggable crRNA, where the presence of the respective ligand would lead to a conformational change that enabled binding of the crRNA to the target sequence. They showed that increasing the concentration of tetracycline or theophylline led to improved transactivation across up to 100 μM and 1000 μM respectively. Finally they also showed that they could employ a G-protein coupled receptor (GPCR) for inducible co-localisation of the TEVp and TCS domain upon ligand binding to the receptor, enabling cleavage and release of dAsCas12a-VPR. They again observed increased transactivation, when employing two different GPCRs, as the ligand concentration was increased.

Finally work by Zhang *et al*., explored the application of the p300 histone acetyltransferase for transactivation (Zhang et al., 2018). They initially screened dAsCas12a and dLbCas12a with a C-terminal fusion of the p300 core domain, targeting with four crRNAs to either the MYOD or IL1RN promoter. However they only observed significant transactivation for dLbCas12a-p300. They proceeded to show that, similarly to dCas9-p300, dLbCas12a-p300 could transactivate downstream genes when targeted to a distal enhancer. They also adapted dLbCas12a to the SunTag system previously described. This system uses a tandem array of peptides, that can be bound by short antibody fragments, fused to the CRISPR scaffold to enable recruitment of effectors bound to the antibody fragments. They showed that by fusing the tandem array of peptides to dLbCas12a they were able to recruit transactivation domains and induce robust transactivation for three different target genes, when providing an individual targeting crRNA.

1.12 Scope of this thesis

Whilst techniques have been developed to enable specific transactivation of targeted genes, several limitations have remained, that provide challenges when applying synthetic transcription factors as tools, as well as limiting potential for therapeutic applications. Here we focus on Cas12a and its adaptation as a synthetic transcription factor. Studies published during the course of this work successfully showed that Cas12a derived synthetic transcription factors could be generated for As and Lb Cas12a. We initially show in Chapter 3 that for the As variant transactivation can be achieved through N-terminal as well as C-terminal transactivation. In Chapter 4 we demonstrate that the FnCas12a variant (possessing a shorter PAM sequence) can be adapted as a synthetic transcription factor and perform a comparison of the As, Fn and Lb synthetic transcription factors, exploring activity and orthogonality. We confirm activity of the dFnCas12a-VPR synthetic transcription factor when independently targeting three endogenous genes in mammalian cells, also observing that increased transactivation is possible when multiple significantly activating crRNAs target a single promoter. In Chapter 5 we demonstrate that short arrays can enable transactivation of target genes by dFnCas12a-VPR and in some cases synergy (greater than the additive transactivation of the individual crRNAs) can be observed. We explore the crRNA arrays, testing up to 9-crRNA arrays expressed from a pol3 promoter and observe that transactivation is achieved, however there is reduced activity for the crRNAs are positioned closer to the 3' of the array. Finally, we go on to show evidence that transactivation can be achieved with a split dFnCas12a-VPR.

Chapter 2 Materials and methods

2.1 Molecular Biology techniques

2.1.1 Oligo annealing

Oligonucleotides were annealed by adding 42 µl of ddH_2O with 5 µl of NEB buffer 3.1 (NEB cat # B7203S), 1.5 ul of 100 µM of the template strand oligonucleotide and 1.5 ul of 100 µM of the non-template strand oligonucleotide to a sterile PCR tube. The reaction was incubated in a thermal cycler with the following protocol. Cycle 1 set at 98ºC for 10 minutes, then 60 subsequent cycles, where the temperature was lowered by 1ºC for 60 cycles. The final step was set at 16ºC held indefinitely.

2.1.2 PNK + Ligase treatment

As crRNA arrays were constructed using ordered oligos, phosphorylation was necessary for successful ligation. For PNK treatment and ligation of inserts to a backbone vector, the following reagents were combined in a PCR tube on ice; 2 µl T4 DNA Ligase buffer (NEB cat #B0202S), 1 µl PNK (NEB cat #M0201S), 1 µl T4 DNA Ligase (NEB cat #M0202), vector DNA, insert DNA and ddH_2O to 20 µl. For generating single crRNA vectors, 50 ng of of Bpi1 digested pU6 vector backbone was added with 1 µl of annealed oligos for the crRNA of interest. The reaction is incubated in a thermal cycler at 37ºC for 30 minutes, then held at 4ºC indefinitely.

2.1.3 Bacterial plates and liquid media

LB media was prepared by adding the following components; 5 g of tryptone (BD diagnostic cat# 211705), 2.5 g of yeast extract (BD diagnostic cat # 212710), 5 g of sodium chloride, 7.5 g of agar (only for plates). These were dissolved in 500 ml of ddH_2O and autoclaved.

LB media requiring antibiotics was made by adding ampicillin from stock solution (100 mg/ml) to a final concentration of 100 μg/ml or adding kanamycin from stock solution (50 mg/ml) to a final concentration of 50 μg/ml.

2.1.4 Chemically competent cell preparation

SOB media was made by mixing the following components; 20 g of tryptone, 5 g of yeast extract, 0.58 g of sodium chloride and 0.19 g of potassium chloride. These were dissolved in 500ml of sterile ddH_2O then autoclaved

The CCMB buffer was prepared by dissolving the following components; 5.9 g of calcium chloride dihydrate (Sigma cat # 223506-500G), 2 g of manganese (II) chloride tetrahydrate (Sigma cat # 221279-100G), 1 g of magnesium chloride hexahydrate (Sigma cat # M9272-100G), 500 ml of 5M potassium acetate (Sigma cat # 95843-100ML-F) and 50 ml of glycerol. All these components were dissolved in 500 ml of ddH_2O then filter sterilised (Sarstedt cat# 83.1823)

The E. coli Mach 1 (F– ϕ80lacZΔM15 ΔlacX74 hsdR(rK–, mK+) ΔrecA1398 endA1 tonA) competent cells was purchased from Thermo fisher (cat # C862003). Cells were streaked out on LB plates without selection and grown overnight in a 37℃ incubator. An overnight liquid culture of these strain was prepared by inoculating 5 ml of LB (without selection) with a colony of the E. coli cells streaked on the LB plate. This was grown overnight in a shaker (Infors HT Multitron Pro) set at 200 rpm at 37℃. The overnight cultures were used to inoculate 500 ml of SOB media and grown in a shaking incubator set at 200 rpm in 37℃ till an OD of 0.4 (approximately 2 hours). On ice in sterile conditions, the cells are pelleted in centrifuge tubes at 3,000 rpm for 2 minutes in 4℃. The supernatant was discarded. The

pellet was resuspended in 100 ml of CCMB then pelleted at 3,000 rpm for 2 minutes at 4℃ and the supernatant was subsequently discarded. The pellet was resuspended in 50 ml of CCMB then pelleted at 3,000 rpm for 2 minutes at 4℃ and the supernatant was subsequently discarded. The pellet was resuspended in 20 ml of CCMB then incubated on ice for 30 minutes to 2 hours. The cells were then aliquoted in 50 µl aliquots, flash frozen on dry ice then stored at -80℃.

2.1.5 *E. coli* transformation by heat shock

Plasmid DNA were transformed into chemically competent E. coli Mach1 cells by heat shock. Competent cells previously prepared (in 50 µl aliquots) were thawed. In sterile conditions, DNA was added to the reaction and kept on ice for 15 minutes. The reaction was transferred to a water bath set at 42℃ for 45 seconds then incubated on ice for 5 minutes. In sterile conditions, 500 µl of SOC was added to the transformation reaction and incubated at 37℃ for 1 hour. The cells are pelleted by centrifugation at 13,000 g for 1 minute and 400 µl of the supernatant was removed. The remaining supernatant was used to resuspend the pellet. The cell suspension was plated on LB with selection (antibiotics) and evenly spread using a disposable plastic spreader. These plates were incubated at 37℃ overnight.

2.1.6 Plasmid extraction from bacterial cultures

5 ml of LB with the appropriate antibiotic selection (100 µg/ml for ampicillin or 50 µg/ml for kanamycin) was aliquoted into 15 ml centrifuge tubes (Sarstedt cat # 62.554.001) and inoculated with bacterial colonies containing the desired plasmids. These cultures were grown overnight in a 37℃ shaker (Infors HT Multitron Pro) set at 200 rpm. Plasmids were extracted using the E.Z.N.A. Plasmid mini kit I (Q-spin) (Omega Biotek cat #D6942-01)

according to the manufacturer's instructions. Plasmids were eluted in 50 µl ddH_2O. The concentration was measured using a nanodrop.

2.1.7 PCR amplification

PCR amplification reactions to construct plasmids were performed using the Q5 DNA polymerase. The PCR reaction was prepared by adding 1x Q5 reaction buffer (5 µl), 0.2 mM dNTPs (1 µl), 0.001 U of Q5 high fidelity polymerase (0.5 µl), 0.5 µM of the forward primer, 0.5 µM of the reverse primer, 1 µl of the template DNA for a final volume of 50 µl. The PCR was incubated in a thermocycler with the following program: 98°C for 30 seconds; 35 cycles of 98°C for 10 seconds, 55°C for 30 seconds and 72°C for 30 seconds; 72°C for 2 minutes; hold at 4°C.

2.1.8 PCR purification

PCR products were purified using the QIAquick PCR purification kit (Qiagen cat #28104) according to the manufacturer's protocol.

2.1.9 Restriction digestion

Digestion of plasmids and PCR products were prepared by mixing 1 µl of plasmid DNA with 1 µl of restrictions enzyme (NEB) and 1x appropriate buffer for the enzyme (NEB) (5 µl) to a final volume of 50 µl. The reaction was incubated at 37°C for 1-2 hours and size verified by agarose gel electrophoresis

2.1.10 Agarose gel electrophoresis

PCR products and digestion reactions were verified by gel electrophoresis. 1% or 2% agarose gel was prepared, depending on the expected size of the band. This was prepared

by weighing the appropriate amount of agarose and measuring the corresponding volume of TAE buffer according to the desired gel percentage. The agarose-TAE mixture was microwaved until the agarose is completely dissolved. The mixture was allowed to cool at 55°C. SYBR safe DNA gel stain (Thermo cat #S33102) was added, mixed thoroughly and poured on a casting tray and allowed to set.

PCR products and digests were mixed with 6x gel loading dye (NEB cat #B7024S) and loaded onto the gel. The agarose gel was run at 125V for 15 minutes in 1xTAE buffer then imaged using a gel documentation system (BioRad Gel Dox XR+)

2.1.11 Gel extraction and purification

Digested plasmids or DNA fragments ran on a gel were separated by size and the desired band was excised using a scalpel and purified using the QIAquick Gel Extraction Kit (Qiagen cat #28704) according to the manufacturer's instructions

2.1.12 Sanger sequencing

Purified plasmids were sent for sequencing, along with the appropriate primers, to the MRC Protein Phosphorylation and Ubiquitylation Unit (PPU) DNA sequencing and services at the University of Dundee.

2.1.13 Gibson assembly

Gibson assembly was carried out using the NEB HiFi Gibson assembly master mix (NEB cat # E2621S). 10 µl of the Gibson assembly master mix reaction was prepared instead and plasmid:insert DNA was added at a 1:1 ratio. The reaction was incubated at 50℃in a

thermocycler for 1 hour and subsequently kept on ice for 5 minutes. The reaction was then transformed into *E. coli* competent cells according to the method in section 2.1.5.

2.1.14 Semi-dry Western Blot

The following transfer buffer was first prepared: 3.03 g/L Tris base, 14.4 g/L glycine, 20 % methanol, Up to 1L dH20. Previously extracted protein samples (resuspended in SDS reducing buffer) were boiled for 5 minutes at 95, before being run in transfer buffer on a 4-15% Mini-PROTEAN® TGX™ Precast Gels alongside NEB Broad-range protein standard. Transfer was performed using a BioRad Mini Trans-Blot® Electrophoretic Transfer Cell, with a PVDF membrane (BIORAD), pre-activated by soaking in Methanol for 15 minutes. After transfer the blots were blocked overnight in 5% milk powder in PBS-T (PBS with 0.1% TRITON-X). The blots were then exposed to the primary antibodies (1:5000 unless otherwise stated) for 1hour, before washing the blot with 5% milk powder in PBS-T and applying the appropriate HRP secondary antibody (1:5000 unless otherwise stated). The western blot was subsequently washed with PBS-T and the blot was imaged using the LI-COR C-DiGit Blot Scanner.

2.2 Mammalian cell culture techniques

2.2.1 HEK293FT media preparation

Media for growing HEK293FT cells was prepared by supplementing Dulbecco's modified Eagle's medium (DMEM) (Thermo Fisher cat #31966021) with 10% Fetal bovine serum (FBS) (Thermo Fisher cat #16000036), 4 mM L-glutamine (cat #) and 1% penicillin streptomycin (Thermo Fisher cat #15140122)

2.2.2 Automated cell counting

In a T75 flask, add 2 ml of TrypLE Express Enzyme (Thermo Fisher cat #12605010) to release the cells from the bottom of the flask. 10 µl of Trypan blue solution (Thermo Fisher cat #15250061) was added to 10 µl of cells. 10 µl of the mixture was pipetted into the Countess Cell Counting Chamber Slides (Thermo Fisher cat #C10228). The slide was inserted into the Countess II Automated Cell Counter (Invitrogen cat #15397802). The cell counting process was started and relevant statistics were recorded.

2.2.3 Freezing down cells (-80℃)

The freezing media was prepared by mixing 50% HEK293FT media, 40% Fetal calf serum (FCS) and 10% DMSO. Split cells were resuspended in freezing media and transferred to cryotubes. They are kept on ice for 30 minutes then wrapped in tissue before transferring them to a -80℃ freezer for long term storage.

2.2.4 Mammalian cell transfection (lipofection)

Cells were seeded 1 day prior to transfection in a 24 well plate then on the day of transfection, a working solution of diluted Lipofectamine was prepared (Thermo Fisher cat #11668027) with the relevant amount of OptiMEM medium (Thermo Fisher cat #31985-047) (25 µl of OptiMEM per 1 µl of Lipofectamine). 25 µl of OptiMEM was aliquoted into eppendorfs, corresponding to each well to be transfected. DNA to be transfected was added to the eppendorf tubes containing OptiMEM. 25 µl of diluted Lipofectamine working solution was added to each eppendorf tube resulting in a total mixture volume of 50 µl. The solution was then gently mixed and incubated at room temperature (20-22℃) for 5 minutes. After incubation, the OptiMEM-DNA-lipofectamine mix was transferred to the cells on the 24 well plate.

2.3 Screening techniques

2.3.1 Luciferase assay

This Luciferase assay utilises the expression of two luciferase genes (Firerfly Luciferase and Renilla Luciferase). Upstream of the minimal promoter driving Firefly Luciferase are multiple binding sites for crRNA or gRNAs to test the activity of their corresponding synthetic transcription factors (dCas12a-VPR for crRNAs and dCas9-VPR for gRNA). The detected luminescence level of Firefly Luciferase was compared to a Renilla luciferase constitutively expressed from a separate plasmid to determine any up regulation in gene expression by the synthetic transcription factor.

Plasmids constructed to contain crRNAs targeting the multiple binding sites upstream of the minimal promoter driving Firefly Luciferase were transfected along with the plasmid containing the dCas12a-VPR synthetic transcription factor. As a positive control, a plasmid containing gRNA targeting the multiple binding sites upstream of the minimal promoter driving Firefly Luciferase was transfected along with the plasmid containing the dCas9-VPR.

These plasmids (gRNA and dCas9-VPR plasmid pair or crRNA and dCas12a-VPR plasmid pair) were transfected along with a Firefly luciferase reporter and a Renilla luciferase internal control plasmid previously constructed at the Rosser lab. After 48 hours post-transfection, the HEK293FT cells were lysed with Passive Lysis buffer. The Firefly and Renilla luciferase induced luminescences for each sample were measured using the Dual Luciferase Reporter Assay System (Promega cat #E1910) using the Modulus II microplate reader (Turner Biosystems)

2.3.2 RNA extraction, Reverse transcription (cDNA synthesis) and Quantitative RT-PCR

Three days post transfection, cells were harvested. The addition of TrypLE Express Enzyme (Thermo Fisher cat #12605010) released the cells from the bottom of the well, pelleted by centrifugation then resuspended in RNA lysis buffer. The RNA was then extracted using the E.Z.N.A Total RNA Kit 1 (Omega Biotek cat #R6834-01) according to the manufacturer's instructions. The concentration and RNA quality was assessed using the nanodrop. 1 µg of RNA was mixed with 1µl of 50µM oligo $d(T)_{20}$ (IDT), 1µl of 10mM dNTP mix (Promega cat #U1240) and dd H_20 up to a final volume of 13µl in a PCR tube and incubated at 65°C for 5 minutes then on ice for 1 minute. The following components were added to each sample; 4µl of SuperScript IV Reverse Transcriptase buffer, 1µl of 0.1 M DTT, 1ul of dH_20, 0.5µl of RiboLock RNase Inhibitor (Thermo Scientific cat #EO0381) and 0.5µl of SuperScript IV Reverse Transcriptase (Thermo Fisher cat #18090050). The reactions were incubated at 52ºC for 10 minutes followed by 80ºC for 10 minutes and holding at 4°C. A qPCR reaction plate was set up with Power SYBR Green qPCR mix (Thermo Fisher cat #4367659), diluted cDNA and primers and ran on the StepOnePlus real-time PCR machine (Thermo Fisher cat #4376600). The results were analysed on the StepOnePlus PCR machine software.

2.3.3 ChIP qPCR

ES cells (provided by Elin Ennervald) were treated with typsin before being spun down and washed with PBS The cells were subsequently resuspended in 18ml PBS and 2ml 10x cross linking buffer. After 10 minutes, rotating at room temperature, the cross-linking reaction was quenched with 2.2ml of Glycine (final concentraltion 125mM. From each plate ~2.5e7 cells were used for H3K27ac ChIP-enrichment.

Chromatin was sheared to a median fragment size of 250 bp using a Bioruptor XL (Diagenode). H3K27ac enrichment was performed by incubation with 5 µg of Abcam (ab4729) and 200 µl of sheep anti-rabbit IgG magnetic beads (Life Technologies 11203D) for 16 hrs at 4°. Cross-links were reversed via overnight incubation at 65°C with sodium dodecyl sulfate, and DNA was purified using MinElute DNA purification columns (Qiagen). 10ng of DNA was used for subsequent qPCR analysis. The ratio of the enriched to 1% INPUT was calculated for positive and negative control conditions to ascertain the fold enrichment at the positive control loci.

2.3.4 Screening for synergy

When screening for synergy, we defined synergy as the combined (pooled or array) experimentally derived distribution being significantly higher than the hypothetical distribution expected from adding the distributions of the composite crRNAs normalised to the negative control. Using this approach we are able to infer the mean of the hypothetical distribution by adding the geometric mean of the fold changes (calculated by taking 2^-CT) and taking the log base 2 of this value to generate the CT distribution for the hypothetical additive condition.

Mean (A+B) = $\log_2$(Geometric mean(Fold change A) + Geometric mean (Fold change B))

We calculated the variance of the mean using standard error propagation; var (f(x,y)) / $(f(x,y))^2$ = (partial derivative f(x,y) / partial derivative x) ^{2} x^2 + (partial derivative f(x,y)/ partial derivative x) ^{2}.

By taking the square root of the variance of the hypothetical condition, we can calculate the standard deviation. With the respective means and standard deviations for the

combined (pool/array) conditions and the hypothetical additive conditions, a Welch's t test can be performed to ascertain the level of significance of separation observed for the two distributions.

2.3.5 Statistical tests

All statistical tests were performed using Prism 8 software. Where levels of significance are displayed on graphs using stars, the stars represent the P values as follows; * = $P < 0.05$, ** = $P < 0.01$, *** = $P < 0.001$.

2.4 Primer list

Table 2.1 - Table of primers used in this study

Primer name	**Primer sequence**
As-OuterF	GGA GAA GGT GCA GCG CAG CC
As-Mut1F	GAC ACC TAT CAT CGG CAT CGC GCG GGG CGA GA
As-Mut1R	CAG GTT TCT CTC GCC CCG CGC GAT GCC GAT G
OuterR	TAG GGA TAA GCG TAA TCT GGA ACA TCG TAT GG
Fn OuterF	TAG GGA GAC CCA AGC TGG CTA GC
Fn Mut1R	TGC CGC TCT CCT CGC GCA ATG CTC AGG ATG TGC
Fn Mut1 F	ACG ATG TGC ACA TCC TGA GCA TTG CGC GAG GAG AGC
Lb Outer F	AGT ATG ACG ATA TCC ACC TGA AGA AGA AGG C
Lb Mut1R	CAG ATT GCG CTC GCC CCT CGC GAT GCC GAT
Lb Mut1F	CCC CTA TGT GAT CGG CAT CGC GAG GGG CGA GC
VPR F	AAAAAGGCCGGCCAGGCAAAAAAGAAAAAGGGATCCGGAACC CGGGCTGACGCATTGGACGATTTTGATCTGG
VPR R	ccctctagactcgagGTCATCCGGAAAACAGAGATGTGTCGAAGAT GGACAGTC
As luc g F	CACCGTAATTTCTACTCTTGTAGATCAGGCTAGCCATGCTTCGCT
As luc g R	AAA CAG CGA AGC ATG GCT AGC CTG ATC TAC AAG AGT AGA AAT TAC
Fn luc g R	CAC CGT AAT TTC TAC TAA GTG TAG ATC AGG CTA GCC ATG CTT CGC T

Fn luc g1 R	AAA CAG CGA AGC ATG GCT AGC CTG ATC TAC ACT TAG TAG AAA TTA C
Lb luc g1 F	CAC CGT AAT TTC TAC TGT TGT AGA TCA GGC TAG CCA TGC TTC GCT
Lb luc g1 R	AAA CAG CGA AGC ATG GCT AGC CTG ATC TAC AAC AGT AGA AAT TAC
As g1 F	CACCGGTAATTTCTACTCTTGTAGATCAAGCTAGCCATGCTTCGC T
As g1 R	AAACAGCGAAGCATGGCTAGCTTGATCTACAAGAGTAGAAATT ACC
As g2 F	CACCGGTAATTTCTACTCTTGTAGATCAGGCTAGTCATGCTTCGC T
As g2 R	AAACAGCGAAGCATGACTAGCCTGATCTACAAGAGTAGAAATT ACC
As g3 F	CACCGGTAATTTCTACTCTTGTAGATCAGGCTAGCGGGGGGCTA TA
As g3 R	AAACTATAGCCCCCCGCTAGCCTGATCTACAAGAGTAGAAATTA CC
Lb g1 F	CACCGGTAATTTCTACTGTTGTAGATCAAGCTAGCCATGCTTCGC T
Lb g1 R	AAACAGCGAAGCATGGCTAGCTTGATCTACAACAGTAGAAATTA CC
Lb g2 F	CACCGGTAATTTCTACTGTTGTAGATCAGGCTAGTCATGCTTCGC T
Lb g2 R	AAACAGCGAAGCATGACTAGCCTGATCTACAACAGTAGAAATTA CC
Lb g3 F	CACCGGTAATTTCTACTGTTGTAGATCAGGCTAGCGGGGGGCTA TA
Lb g3 R	AAACTATAGCCCCCCGCTAGCCTGATCTACAACAGTAGAAATTA CC
F primer VPR	TTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGCGTTTA AACTTAAGCTTGGTACGCCACCATGGACGCATTGGACGATTTTG ATCTGG
R VPR AsCpf1	CTCAAACCGCAGTGTCTTGCTCACCTGATACAGGTTGGTAAAGC CCTCGAACTGTGTCATAGCCCGTCCGGAACCGCTGGCCTCTTTG TAGCCTGCTTTTTTGTACAAACTTGTTGATATCAACGCGTCAAGT CGACGGATCCCTGGCGAAAACAGAGATGTGTCGAAGATGGACA GTC
N term As F	ATGACACAGTTCGAGGGCTTTACCAACCTGTATCAGGTGAGCAA GACACT
N term As R	TCGCTGGTTTTCTGCTTGAAGGCCTCGCTCAGCTCC
As PmlI F	GCG ACA AGT TCT TTT TCC AC
As Site 1 PmlI R	CAC GTC GTA GTG CAG AAA GT
As Site 1 BamHI F	AAA ACC GGC GAC TTC ATC CT
As BamHI R	CTG GAA CAT CGT ATG GGT AG
As Site 2 BamHI F	GAT GGC TCC AAC ATC CTG CC

As Site 2 PmlI R	CCT GAA CAC GAT GCC CTT CT
VP64 Site1 5AA F	ACT TTC TGC ACT ACG ACG TGG GTG GCT CTG GAG GTG ACG CAT TGG ACG ATT TTG
VP64 Site1 5AA R	AGG ATG AAG TCG CCG GTT TTT CCT CCA CTA CCG CCC AGC ATG TCC AGG TCG
VP64 Site2 5AA F	AGA AGG GCA TCG TGT TCA GGG GTG GCT CTG GAG GTG ACG CAT TGG ACG ATT TTG
VP64 Site2 5AA R	GGC AGG ATG TTG GAG CCA TCT CCT CCA CTA CCG CCC AGC ATG TCC AGG TCG
Outer Alaz As F	CACGAGGATATCAACCTGCAGGAGATCATC
Inner Alaz As R	CTCTGTGGGCTCGAAGCTCAGG
VP64 Alaz As F	TGAGCTTCGAGCCCACAGAGGACGCATTGGACGATTTTGATCTG GATATGCT
VP64 Alaz As R	TCAAAGCCCTCGCTGGTTTTCAGCATGTCCAGGTCGAAATCATC AAG
Inner Alaz As F	AAAACCAGCGAGGGCTTTGATAAGATGTACTATGAC
Outer Alaz As R	TAGTTCAGTGTGATAGGCACGTGGAAAAAGAACTTG
Alt PAM As Luc F	CACCGGTAATTTCTACTCTTGTAGATAGGCTAGCCATGCTTCGCT A
Alt PAM As Luc R	AAACTAGCGAAGCATGGCTAGCCTATCTACAAGAGTAGAAATTA CC
Alt PAM Fn Luc F	CACCGGTAATTTCTACTGTTGTAGATAGGCTAGCCATGCTTCGCT A
Alt PAM Fn Luc R	AAACTAGCGAAGCATGGCTAGCCTATCTACAACAGTAGAAATTA CC
Alt PAM Lb Luc F	CACCGAATTTCTACTAAGTGTAGATAGGCTAGCCATGCTTCGCT A
Alt PAM Lb Luc R	aaacTAGCGAAGCATGGCTAGCCTATCTACACTTAGTAGAAATTC
First AR MST F	CACCGAATTTCTACTAAGTGTAGATAGAGTC
First AR MST R	CTCATCCAGACTCTATCTACACTTAGTAGAAATTC
AR MST 2 F	TGGATGAGAAATGCAATTTCTACTAAGTGTAGATTACCCT
AR MST 2 R	CAGAGAAGAGGGTAATCTACACTTAGTAGAAATTGCATTT
AR MST 3 F	CTTCTCTGCCTTTCAATTTCTACTAAGTGTAGATCTCTAG
AR MST 3 R	GAGGGTTCCTAGAGATCTACACTTAGTAGAAATTGAAAGG
AR MST Last F	GAACCCTCAGCCCCAATTTCTACTAAGTGTAGAT
AR MST Last R	AAACATCTACACTTAGTAGAAATTGGGGCT
HBB MST First F	CACCGAATTTCTACTAAGTGTAGATTACTGA
HBB MST First R	CCATACCATCAGTAATCTACACTTAGTAGAAATTC
HBB MST 2 F	TGGTATGGGGCCAAAATTTCTACTAAGTGTAGATAAGTCC

HBB MST 2 R	TAGGAGTTGGACTTATCTACACTTAGTAGAAATTTTGGCC
HBB MST 3 F	AACTCCTAAGCCAGAATTTCTACTAAGTGTAGATCAAGTG
HBB MST 3 R	CGTAAATACACTTGATCTACACTTAGTAGAAATTCTGGCT
HBB MST Last F	TATTTACGTAATATAATTTCTACTAAGTGTAGAT
HBB MST Last R	AAACATCTACACTTAGTAGAAATTATATTA
NPY1R MST First F	CACCGAATTTCTACTAAGTGTAGATAAGCCT
NPY1R MST First R	GTTTCCCGAGGCTTATCTACACTTAGTAGAAATTC
NPY1R MST 2 F	CGGGAAACTGCCCTAATTTCTACTAAGTGTAGATTTTGTT
NPY1R MST 2 R	GACCTGCAAACAAAATCTACACTTAGTAGAAATTAGGGCA
NPY1R MST 3 F	TGCAGGTCAGTGCCAATTTCTACTAAGTGTAGATGGCTGG
NPY1R MST 3 R	CTCGAGCGCCAGCCATCTACACTTAGTAGAAATTGGCACT
NPY1R MST Last F	CGCTCGAGCTCTCCAATTTCTACTAAGTGTAGAT
NPY1R MST Last R	AAACATCTACACTTAGTAGAAATTGGAGAG
Fn DR F	CacCGTAATTTCTACTGTTGTAGATgCgtcttcgaattcgaagacct
Fn DR R	aaacaggtcttcgaattcgaagacGcATCTACAACAGTAGAAATTAC
ASCL1 c1 F	AGATAGCTGGGTTTGTTGTTGCAG
ASCL1 c1 R	AAACCTGCAACAACAAACCCAGCT
ASCL1 c2 F	AGATCAAGGAGcgggagaaaggaa
ASCL1 c2 R	AAACttcctttctcccgCTCCTTG
ASCL1 c3 F	AGATgggagtgggtgggaggaaga
ASCL1 c3 R	AAACtcttcctcccacccactccc
IL1RN c1 F	AGATCGCAGATAAGAACCAGTTTG
IL1RN c1 R	AAACCAAACTGGTTCTTATCTGCG
IL1RN c2 F	AGATCAGGAGGGTGACTCAGGCTA
IL1RN c2 R	AAACTAGCCTGAGTCACCCTCCTG
IL1RN c3 F	AGATGCATCAAGTCAGCCATCAGC
IL1RN c3 R	AAACGCTGATGGCTGACTTGATGC
Fn HBB g1 F 2	CACCGTAATTTCTACTGTTGTAGATTACTGATGGTATGGGGCCAA
Fn HBB g1 R 2	AAACTTGGCCCCATACCATCAGTAATCTACAACAGTAGAAATTAC

Fn HBB g2 F 2	CACCGTAATTTCTACTGTTGTAGATAAGTCCAACTCCTAAGCCAG
Fn HBB g2 R 2	AAACCTGGCTTAGGAGTTGGACTTATCTACAACAGTAGAAATTAC
Fn HBB g3 F 2	CACCGTAATTTCTACTGTTGTAGATCAAGTGTATTTACGTAATAT
Fn HBB g3 R 2	AAACATATTACGTAAATACACTTGATCTACAACAGTAGAAATTAC
ASCL1 crRNA 4 F	AGATTTGTTGCAGTGCGTGCGCCT
ASCL1 crRNA 4 R	AAACAGGCGCACGCACTGCAACAA
ASCL1 crRNA 5 F	AGATtcccgCTCCTTGCAAACTCT
ASCL1 crRNA 5 R	AAACAGAGTTTGCAAGGAGcggga
ASCL1 crRNA 6 F	AGATctttctcccgCTCCTTGCAA
ASCL1 crRNA 6 R	AAACTTGCAAGGAGcgggagaaag
HBB crRNA 4 F	AGATGTAGCAATTTGTACTGATGG
HBB crRNA 4 R	AAACCCATCAGTACAAATTGCTAC
HBB crRNA 5 F	AGATGAGGGAGGGCTGAGGGTTTG
HBB crRNA 5 R	AAACCAAACCCTCAGCCCTCCCTC
IL1RN crRNA 4 F	AGATTCTGCATGTGACCTCCCATC
IL1RN crRNA 4 R	AAACGATGGGAGGTCACATGCAGA
IL1RN crRNA 5 F	AGATGTTTCTGCTAGCCTGAGTCA
IL1RN crRNA 5 R	AAACTGACTCAGGCTAGCAGAAAC
IL1RN crRNA 6 F	AGATGCCAGCATGAGGAGATGGGC
IL1RN crRNA 6 R	AAACGCCCATCTCCTCATGCTGGC
A 2X 1F	CACCGTAATTTCTACTGTTGTAGATCAAGGA
A 2X 1R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAC
A 2X 2F	GcgggagaaaggaaTAATTTCTACTGTTGTAGATTTGTTG
A 2X 2R	ACGCACTGCAACAAATCTACAACAGTAGAAATTAttcctt
A 2X 3F	CAGTGCGTGCGCCTTAATTTCTACTGTTGTAGAT
A 2X 3R	aaacATCTACAACAGTAGAAATTAAGGCGC
H1 H2 2X 1F	CACCGTAATTTCTACTGTTGTAGATTACTGA
H1 H2 2X 1R	CCATACCATCAGTAATCTACAACAGTAGAAATTAC
H1 H2 2X 2F	TGGTATGGGGCCAATAATTTCTACTGTTGTAGATAAGTCC

H1 H2 2X 2R	TAGGAGTTGGACTTATCTACAACAGTAGAAATTATTGGCC
H1 H2 2X 3F	AACTCCTAAGCCAGTAATTTCTACTGTTGTAGAT
H1 H2 2X 3R	aaacATCTACAACAGTAGAAATTACTGGCT
H4 H5 2X 1F	CACCGTAATTTCTACTGTTGTAGATGTAGCA
H4 H5 2X 1R	GTACAAATTGCTACATCTACAACAGTAGAAATTAC
H4 H5 2X 2F	ATTTGTACTGATGGTAATTTCTACTGTTGTAGATGAGGGA
H4 H5 2X 2R	CTCAGCCCTCCCTCATCTACAACAGTAGAAATTACCATCA
H4 H5 2X 3F	GGGCTGAGGGTTTGTAATTTCTACTGTTGTAGAT
H4 H5 2X 3R	aaacATCTACAACAGTAGAAATTACAAACC
I4 I6 2X 1F	CACCGTAATTTCTACTGTTGTAGATTCTGCA
I4 I6 2X 1R	AGGTCACATGCAGAATCTACAACAGTAGAAATTAC
I4 I6 2X 2F	TGTGACCTCCCATCTAATTTCTACTGTTGTAGATGCCAGC
I4 I6 2X 2R	CTCCTCATGCTGGCATCTACAACAGTAGAAATTAGATGGG
I4 I6 2X 3F	ATGAGGAGATGGGCTAATTTCTACTGTTGTAGAT
I4 I6 2X 3R	aaacATCTACAACAGTAGAAATTAGCCCAT
A2 first F	CACCGTAATTTCTACTGTTGTAGATCAAGGA
A24 F	GcgggagaaaggaaTAATTTCTACTGTTGTAGATTTGTTG
A4H1 F	CAGTGCGTGCGCCTTAATTTCTACTGTTGTAGATTACTGA
H1H2 F	TGGTATGGGGCCAATAATTTCTACTGTTGTAGATAAGTCC
H2I4 F	AACTCCTAAGCCAGTAATTTCTACTGTTGTAGATTCTGCA
I4I6 F	TGTGACCTCCCATCTAATTTCTACTGTTGTAGATGCCAGC
I6 last F	ATGAGGAGATGGGCTAATTTCTACTGTTGTAGAT
A4I4 F	CAGTGCGTGCGCCTTAATTTCTACTGTTGTAGATTCTGCA
I6H1 F	ATGAGGAGATGGGCTAATTTCTACTGTTGTAGATTACTGA
H2 last F	AACTCCTAAGCCAGTAATTTCTACTGTTGTAGAT
H1 first F	CACCGTAATTTCTACTGTTGTAGATTACTGA
I6A2 F	ATGAGGAGATGGGCTAATTTCTACTGTTGTAGATCAAGGA
A4 Last F	CAGTGCGTGCGCCTTAATTTCTACTGTTGTAGAT
H2A2 F	AACTCCTAAGCCAGTAATTTCTACTGTTGTAGATCAAGGA
I4 first F	CACCGTAATTTCTACTGTTGTAGATTCTGCA

A2 first R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAC
A24 R	ACGCACTGCAACAAATCTACAACAGTAGAAATTAttcctt
A4H1 R	CCATACCATCAGTAATCTACAACAGTAGAAATTAAGGCGC
H1H2 R	TAGGAGTTGGACTTATCTACAACAGTAGAAATTATTGGCC
H2I4 R	AGGTCACATGCAGAATCTACAACAGTAGAAATTACTGGCT
I4I6 R	CTCCTCATGCTGGCATCTACAACAGTAGAAATTAGATGGG
I6 last R	aaacATCTACAACAGTAGAAATTAGCCCAT
A4I4 R	AGGTCACATGCAGAATCTACAACAGTAGAAATTAAGGCGC
I6H1 R	CCATACCATCAGTAATCTACAACAGTAGAAATTAGCCCAT
H2 last R	aaacATCTACAACAGTAGAAATTACTGGCT
H1 first R	CCATACCATCAGTAATCTACAACAGTAGAAATTAC
I6A2 R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAGCCCAT
A4 Last R	aaacATCTACAACAGTAGAAATTAAGGCGC
H2A2 R	tctcccgCTCCTTGATCTACAACAGTAGAAATTACTGGCT
I4 first R	AGGTCACATGCAGAATCTACAACAGTAGAAATTAC
NT 1st F	CacCGTAATTTCTACTGTTGTAGATcgacgc
NT 1st R	acccggcggcgtcgATCTACAACAGTAGAAATTAC
NT 12 F	cgccgggtcacaacTAATTTCTACTGTTGTAGATaaacgc
NT 12 R	ttggttcggcgtttATCTACAACAGTAGAAATTAgttgtg
NT 23 F	cgaaccaactcgcgTAATTTCTACTGTTGTAGATgcgagc
NT 23 R	atgggtgcgctcgcATCTACAACAGTAGAAATTAcgcgag
NT 34 F	gcacccattggaccTAATTTCTACTGTTGTAGATgcgtcc
NT 34 R	ggtcttgcggacgcATCTACAACAGTAGAAATTAggtcca
NT 45 F	gcaagacctatccaTAATTTCTACTGTTGTAGATccgtcg
NT 45 R	ccgcgaaccgacggATCTACAACAGTAGAAATTAtggata
NT 56 F	gttcgcggctgcagTAATTTCTACTGTTGTAGATatttat
NT 56 R	gcgtgccgataaatATCTACAACAGTAGAAATTActgcag
NT last F	cggcacgccgcaatTAATTTCTACTGTTGTAGAT
NT last R	aaacATCTACAACAGTAGAAATTAattgcg
A 1st 1 F	CacCGTAATTTCTACTGTTGTAGATCAAGGA

A 1st 1 R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAC
A 1st 2 F	GcgggagaaaggaaTAATTTCTACTGTTGTAGATaaacgc
A 1st 2 R	ttggttcggcgtttATCTACAACAGTAGAAATTAttcctt
A 2nd 1 F	cgccgggtcacaacTAATTTCTACTGTTGTAGATCAAGGA
A 2nd 1 R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAgttgtg
A 2nd 2 F	GcgggagaaaggaaTAATTTCTACTGTTGTAGATgcgagc
A 2nd 2 R	atgggtgcgctcgcATCTACAACAGTAGAAATTAttcctt
A 3rd 1 F	cgaaccaactcgcgTAATTTCTACTGTTGTAGATCAAGGA
A 3rd 1 R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAcgcgag
A 3rd 2 F	GcgggagaaaggaaTAATTTCTACTGTTGTAGATgcgtcc
A 3rd 2 R	ggtcttgcggacgcATCTACAACAGTAGAAATTAttcctt
A 4th 1 F	gcacccattggaccTAATTTCTACTGTTGTAGATCAAGGA
A 4th 1 R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAggtcca
A 4th 2 F	GcgggagaaaggaaTAATTTCTACTGTTGTAGATccgtcg
A 4th 2 R	ccgcgaaccgacggATCTACAACAGTAGAAATTAttcctt
A 5th 1 F	gcaagacctatccaTAATTTCTACTGTTGTAGATCAAGGA
A 5th 1 R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAtggata
A 5th 2 F	GcgggagaaaggaaTAATTTCTACTGTTGTAGATatttat
A 5th 2 R	gcgtgccgataaatATCTACAACAGTAGAAATTAttcctt
A 6th 1 F	gttcgcggctgcagTAATTTCTACTGTTGTAGATCAAGGA
A 6th 1 R	tctcccgCTCCTTGATCTACAACAGTAGAAATTActgcag
A 6th 2 F	GcgggagaaaggaaTAATTTCTACTGTTGTAGAT
A 6th 2 R	aaacATCTACAACAGTAGAAATTAttcctt
Fn 1 A1 orig F	CACCGTAATTTCTACTGTTGTAGATCAAGGA
Fn 1 A1 orig R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAC
Fn 2 A1 orig F	GcgggagaaaggaaTAATTTCTACTGTTGTAGATcgagcc
Fn 2 A1 orig R	aaggttatggctcgATCTACAACAGTAGAAATTAttcctt
Fn 3 1-2 orig F	ataaccttaggtgtTAATTTCTACTGTTGTAGATtgactt
Fn 3 1-2 orig R	gcggcttcaagtcaATCTACAACAGTAGAAATTAacacct
Fn 4 2-3 orig F	gaagccgcaacgttTAATTTCTACTGTTGTAGATtgcagc

Fn 4 2-3 orig R	acggacgtgctgcaATCTACAACAGTAGAAATTAaacgtt
Fn 5 H5 orig F	acgtccgtattgaaTAATTTCTACTGTTGTAGATTACTGA
Fn 5 H5 orig R	CCATACCATCAGTAATCTACAACAGTAGAAATTAttcaat
Fn 6 H5 orig F	TGGTATGGGGCCAATAATTTCTACTGTTGTAGATaggccg
Fn 6 H5 orig R	cgaaagctcggcctATCTACAACAGTAGAAATTATTGGCC
Fn 7 4-5 orig F	agctttcgaatcatTAATTTCTACTGTTGTAGATcaactg
Fn 7 4-5 orig R	acagcgcacagttgATCTACAACAGTAGAAATTAatgatt
Fn 8 5-6 orig F	tgcgctgtaagtcaTAATTTCTACTGTTGTAGATgacaga
Fn 8 5-6 orig R	tagtgcaatctgtcATCTACAACAGTAGAAATTAtgactt
Fn 9 I9 orig F	ttgcactagtcgtcTAATTTCTACTGTTGTAGATCAGGAG
Fn 9 I9 orig R	GAGTCACCCTCCTGATCTACAACAGTAGAAATTAgacgac
Fn 10 I9 orig F	GGTGACTCAGGCTATAATTTCTACTGTTGTAGAT
Fn 10 I9 orig R	aaacATCTACAACAGTAGAAATTATAGCCT
Fn 5 I5 orig F	acgtccgtattgaaTAATTTCTACTGTTGTAGATCAGGAG
Fn 5 I5 orig R	GAGTCACCCTCCTGATCTACAACAGTAGAAATTAttcaat
Fn 6 I5 orig F	GGTGACTCAGGCTATAATTTCTACTGTTGTAGATaggccg
Fn 6 I5 orig R	cgaaagctcggcctATCTACAACAGTAGAAATTATAGCCT
Fn 9 H9 orig F	ttgcactagtcgtcTAATTTCTACTGTTGTAGATTACTGA
Fn 9 H9 orig R	CCATACCATCAGTAATCTACAACAGTAGAAATTAgacgac
Fn 10 H9 orig F	TGGTATGGGGCCAATAATTTCTACTGTTGTAGAT
Fn 10 H9 orig R	aaacATCTACAACAGTAGAAATTATTGGCC
Fn 1 H1 orig F	CACCGTAATTTCTACTGTTGTAGATTACTGA
Fn 1 H1 orig R	CCATACCATCAGTAATCTACAACAGTAGAAATTAC
Fn 2 H1 orig F	TGGTATGGGGCCAATAATTTCTACTGTTGTAGATcgagcc
Fn 2 H1 orig R	aaggttatggctcgATCTACAACAGTAGAAATTATTGGCC
Fn 5 A5 orig F	acgtccgtattgaaTAATTTCTACTGTTGTAGATCAAGGA
Fn 5 A5 orig R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAttcaat
Fn 6 A5 orig F	GcgggagaaaggaaTAATTTCTACTGTTGTAGATaggccg
Fn 6 A5 orig R	cgaaagctcggcctATCTACAACAGTAGAAATTAttcctt
Fn 9 A9 orig F	ttgcactagtcgtcTAATTTCTACTGTTGTAGATCAAGGA

Fn 9 A9 orig R	tctcccgCTCCTTGATCTACAACAGTAGAAATTAgacgac
Fn 10 A9 orig F	GcgggagaaaggaaTAATTTCTACTGTTGTAGAT
Fn 10 A9 orig R	aaacATCTACAACAGTAGAAATTAttcctt
Fn 1 I1 orig F	CACCGTAATTTCTACTGTTGTAGATCAGGAG
Fn 1 I1 orig R	GAGTCACCCTCCTGATCTACAACAGTAGAAATTAC
Fn 2 I1 orig F	GGTGACTCAGGCTATAATTTCTACTGTTGTAGATcgagcc
Fn 2 I1 orig R	aaggttatggctcgATCTACAACAGTAGAAATTATAGCCT
Fn 1 A1 Mut F	CACCGTAATTCCTACTGTTGTAGGTCAAGGA
Fn 1 A1 Mut R	tctcccgCTCCTTGACCTACAACAGTAGGAATTAC
Fn 2 A1 Mut F	GcgggagaaaggaaTAATTCCTACTGTTGTAGGTcgagcc
Fn 2 A1 Mut R	aaggttatggctcgACCTACAACAGTAGGAATTAttcctt
Fn 3 1-2 Mut F	ataaccttaggtgtTAATTCCTACTGTTGTAGGTtgactt
Fn 3 1-2 Mut R	gcggcttcaagtcaACCTACAACAGTAGGAATTAacacct
Fn 4 2-3 Mut F	gaagccgcaacgttTAATTCCTACTGTTGTAGGTtgcagc
Fn 4 2-3 Mut R	acggacgtgctgcaACCTACAACAGTAGGAATTAaacgtt
Fn 5 H5 Mut F	acgtccgtattgaaTAATTCCTACTGTTGTAGGTTACTGA
Fn 5 H5 Mut R	CCATACCATCAGTAACCTACAACAGTAGGAATTAttcaat
Fn 6 H5 Mut F	TGGTATGGGGCCAATAATTCCTACTGTTGTAGGTaggccg
Fn 6 H5 Mut R	cgaaagctcggcctACCTACAACAGTAGGAATTATTGGCC
Fn 7 4-5 Mut F	agctttcgaatcatTAATTCCTACTGTTGTAGGTcaactg
Fn 7 4-5 Mut R	acagcgcacagttgACCTACAACAGTAGGAATTAatgatt
Fn 8 5-6 Mut F	tgcgctgtaagtcaTAATTCCTACTGTTGTAGGTgacaga
Fn 8 5-6 Mut R	tagtgcaatctgtcACCTACAACAGTAGGAATTAtgactt
Fn 9 I9 Mut F	ttgcactagtcgtcTAATTCCTACTGTTGTAGGTCAGGAG
Fn 9 I9 Mut R	GAGTCACCCTCCTGACCTACAACAGTAGGAATTAgacgac
Fn 10 I9 Mut F	GGTGACTCAGGCTATAATTCCTACTGTTGTAGGT
Fn 10 I9 Mut R	aaacACCTACAACAGTAGGAATTATAGCCT
Fn 5 I5 Mut F	acgtccgtattgaaTAATTCCTACTGTTGTAGGTCAGGAG
Fn 5 I5 Mut R	GAGTCACCCTCCTGACCTACAACAGTAGGAATTAttcaat
Fn 6 I5 Mut F	GGTGACTCAGGCTATAATTCCTACTGTTGTAGGTaggccg

Fn 6 I5 Mut R	cgaaagctcggcctACCTACAACAGTAGGAATTATAGCCT
Fn 9 H9 Mut F	ttgcactagtcgtcTAATTCCTACTGTTGTAGGTTACTGA
Fn 9 H9 Mut R	CCATACCATCAGTAACCTACAACAGTAGGAATTAgacgac
Fn 10 H9 Mut F	TGGTATGGGGCCAATAATTCCTACTGTTGTAGGT
Fn 10 H9 Mut R	aaacACCTACAACAGTAGGAATTATTGGCC
Fn 1 H1 Mut F	CACCGTAATTCCTACTGTTGTAGGTTACTGA
Fn 1 H1 Mut R	CCATACCATCAGTAACCTACAACAGTAGGAATTAC
Fn 2 H1 Mut F	TGGTATGGGGCCAATAATTCCTACTGTTGTAGGTcgagcc
Fn 2 H1 Mut R	aaggttatggctcgACCTACAACAGTAGGAATTATTGGCC
Fn 5 A5 Mut F	acgtccgtattgaaTAATTCCTACTGTTGTAGGTCAAGGA
Fn 5 A5 Mut R	tctcccgCTCCTTGACCTACAACAGTAGGAATTAttcaat
Fn 6 A5 Mut F	GcgggagaaaggaaTAATTCCTACTGTTGTAGGTaggccg
Fn 6 A5 Mut R	cgaaagctcggcctACCTACAACAGTAGGAATTAttcctt
Fn 9 A9 Mut F	ttgcactagtcgtcTAATTCCTACTGTTGTAGGTCAAGGA
Fn 9 A9 Mut R	tctcccgCTCCTTGACCTACAACAGTAGGAATTAgacgac
Fn 10 A9 Mut F	GcgggagaaaggaaTAATTCCTACTGTTGTAGGT
Fn 10 A9 Mut R	aaacACCTACAACAGTAGGAATTAttcctt
Fn 1 I1 Mut F	CACCGTAATTCCTACTGTTGTAGGTCAGGAG
Fn 1 I1 Mut R	GAGTCACCCTCCTGACCTACAACAGTAGGAATTAC
Fn 2 I1 Mut F	GGTGACTCAGGCTATAATTCCTACTGTTGTAGGTcgagcc
Fn 2 I1 Mut R	aaggttatggctcgACCTACAACAGTAGGAATTATAGCCT
SETD2 F	GCGTTGATATCAACAAGTTTACAGGGAGAGACATCAGTGCCCC
SETD2 R	TAGATGATATCAACCACTTTCTACTTACGAGATCGTTCCTTCTTC ATTTTCCCTCC
SETD1A F	GCGTTGATATCAACAAGTTTACCCCCGCCGCCAC
SETD1A F	TAGATGATATCAACCACTTTtcagtttagggagccccgg

2.5 Plasmids list

Table 2.2 - List of plasmids used in this study

Name of plasmid	Source
Luc gRNA	(Kleinjan et al.)
pTK Renilla	(Kleinjan et al.)
p8BSI	(Kleinjan et al.)
U6 no scaf	(Kleinjan et al.)
dCas9-VPR	(Chavez et al.)
AsCas12a	(Zetsche et al.)
FnCas12a	(Zetsche et al.)
LbCas12a	(Zetsche et al.)
dAsCas12a-VPR	Cloned
dFnCas12a-VPR	Cloned
dLbCas12a-VPR	Cloned
VPR-dAsCas12a	Cloned
dAsCas12a VP64 Site 1 5AA linker	Cloned
dAsCas12a VP64 Site 2 5AA linker	Cloned
dCas9-SETD2	Cloned
dCas9 SETD1A	Cloned

2.6 qPCR primers

Table 2.3 - List of qPCR primers used in this study

qPCR primers	Column1	Source
ASCL1 qPCR F	CGCGGCCAACAAGAAGATG	Kleinjan et al. 2017
ASCL1 qPCR R	CGACGAGTAGGATGAGACCG	Kleinjan et al. 2017
HBB qPCR F	AAG CTG CAC GTG GAT CCT GA	Designed and calibrated
HBB qPCR R	ATT AGC CAC ACC AGC CAC CA	Designed and calibrated
IL1RN qPCR F	GGAATCCATGGAGGGAAGAT	Kleinjan et al. 2017
IL1RN qPCR R	TGTTCTCGCTCAGGTCAGTG	Kleinjan et al. 2017
NPY1R qPCR F	CCATCGGACTCTCATAGGTTGTC	Tak et al. 2017
NPY1R qPCR R	GACCTGTACTTATTGTCTCTCATC	Tak et al. 2017
AR qPCR F	ATGGTGAGCAGAGTGCCCTATC	Tak et al. 2017
AR qPCR R	ATGGTCCCTGGCAGTCTCCAAA	Tak et al. 2017
X5 qPCR F	tcatgtgacctgccctctagt	Buonomo lab
X5 qPCR R	caccctaccataatgcacca	Buonomo lab
TSIX qPCR F	GACTTCTCTGCCCGTGAAAC	Buonomo lab
TSIX qPCR R	CACCCTTCTCCTCTCTGCAC	Buonomo lab

Chapter 3 Generating Cas12a derived synthetic transcription factors

3.1 Introduction

The initial focus of the project was the generation of synthetic transcription factors, taking some of the advancements achieved using the Cas9 (Figure 3.1A) and translating these to Cas12a (Figure 3.1B). As previously described in Chapter 1, key advances have been made in the field of synthetic transcription factors, with adaptation of the Cas9 platform making transactivation of target genes highly tractable. Cas9 based synthetic transcription factors only require the generation of a unique gRNA for targeting a specific locus, in contrast to protein engineering required for zinc finger or TALE based approaches. Such synthetic Cas9-based transcription factors were obtained by first generating DNase inactive variants of Cas9 (dCas9) through single amino acid substitutions in the RuvC and HNH domains (Jinek *et al*., 2012), followed by the fusion of transactivation domains such as VP64 to dCas9 to allow up-regulation of targeted genes (Maeder et al., 2013a).

Here, we aim to build on this work by adapting more recently discovered CRISPR variants which possess interesting and improved properties compared to CRISPR/Cas9. In particular, Cas12a/Cpf1 has been demonstrated to possess the ability to process it's own crRNA arrays (Fonfara et al., 2016), and does not require tracrRNA for the targeting of loci (Zetsche et al., 2015a). This feature of Cas12a facilitates the generation of multiple unique targeting crRNAs, with transcripts

being potentially able to up-regulate multiple target genes simultaneously. This in turn allows more loci to be targeted within a short length of DNA, when compared to Cas9 systems. This is of particular use in applications where size limitations present a barrier, such as the delivery of genetic material using the adeno-associated virus. Finally, the ability to express multiple crRNAs from a single transcript presents applications focused on control and regulation, where it is desirable to have two or more actions directly coupled to one another, such as simultaneous up-regulation of multiple genes.

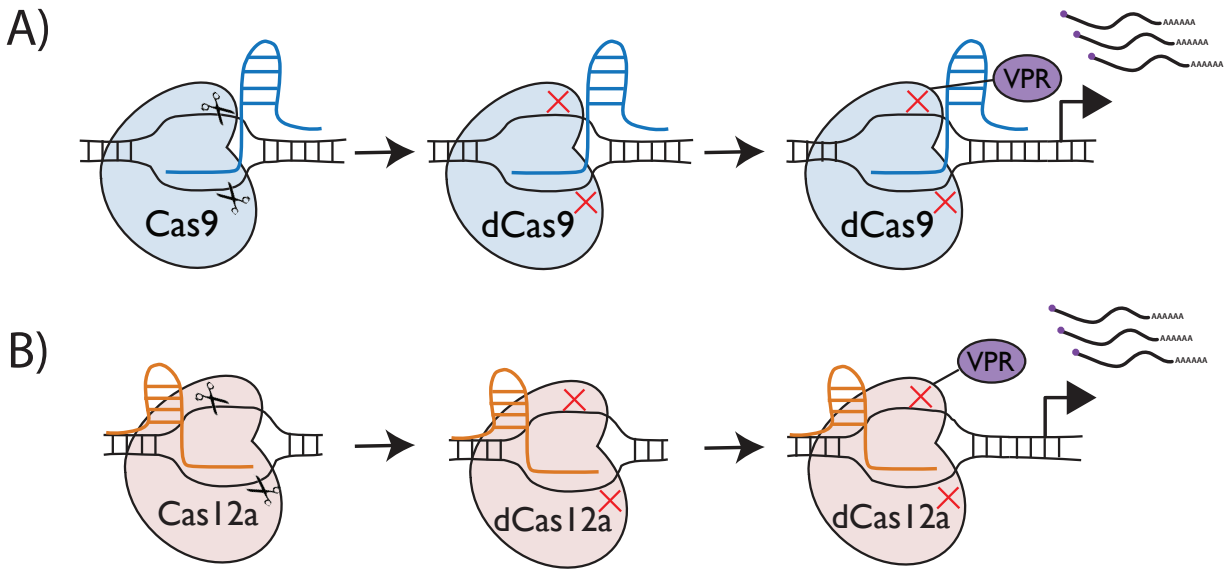

Figure 3-1 - Diagrammatic representation of initial aims of Chapter 3

Diagrammatic representation of the adaptation of (A) CRISPR/Cas9 and (B) Cas12a. First, amino acid substitutions enable inactivation of DNase activity and next effector domains such as the transactivation domain VPR can be fused to the catalytically inactive dCas9 to enable generation of a synthetic transcription factor.

3.2 Generation of dCas12a

We generated DNase inactive variants of Cas12a derived from three different species. These Cas12a variants were chosen based on their clearly defined and relatively short PAM sequences, and their relatively high likelihood of binding DNA/chromatin in a mammalian cell context. Two of these variants, AsCas12a (derived from *Acidaminococcus sp*.) and LbCas12a (derived from *Lachnospiraceae bacterium*) were highlighted in the original paper by Zetsche *et al*. as having cleaved the human genome (Zetsche et al., 2015b). It was therefore hypothesised that, like Cas9, they had a high likelihood of being able to bind DNA when the DNase activity was abolished. Both the AsCas12a and LbCas12a variants were characterised as having a PAM sequence of 'TTTN'. The third variant, FnCas12a (derived from *Francisella tularensis subsp. novicida*) was selected because it was characterised to have the shortest PAM sequence in the family 'TTN' as opposed to 'TTTN', and also showcased faint bands indicating cleavage of the human genome in cleavage assays reported by Zetsche *et al*.. Addgene plasmids for the mammalian codon-optimised variants of AsCas12a (#69982), FnCas12a (#69976) and LbCas12a (#69988) were a gift from Feng Zhang (Table 2.2). The plasmids were modified to generate DNase inactive variants of Cas12a, through a conserved amino acid substitution (Figure 3.2B), previously reported to abolish cleavage (Zetsche et al., 2015b). We generated these amino acid substitutions corresponding to D908A for AsCas12a, D917A for FnCas12a, and D832A for LbCas12a using multichange isothermal mutagenesis (Mitchell *et al*., 2013) (Figure 3.2C). Primers were designed to enable amplification of Cas12a in two parts using the following primer pairs; As-OuterF + As Mut1R and As Mut1F + OuterR for the As variant, Fn-OuterF + Fn Mut1R and Fn Mut1F +

OuterR for the Fn variant, Lb-OuterF + Lb Mut1R and Lb Mut1F + OuterR for the Lb variant (Table 2.1). The primers were designed such that the amino acid substitution is incorporated by the reverse primer for the first amplicon and the forward primer for the second amplicon (Figure 3.2C). After PCR purification of these amplicons using a QIAquick PCR Purification Kit and digestion of the original plasmids with SbfI and BamHI to remove the original coding sequence, the amplicons were assembled with their respective linearised backbone, thus incorporating the mutation for each Cas12a variant.

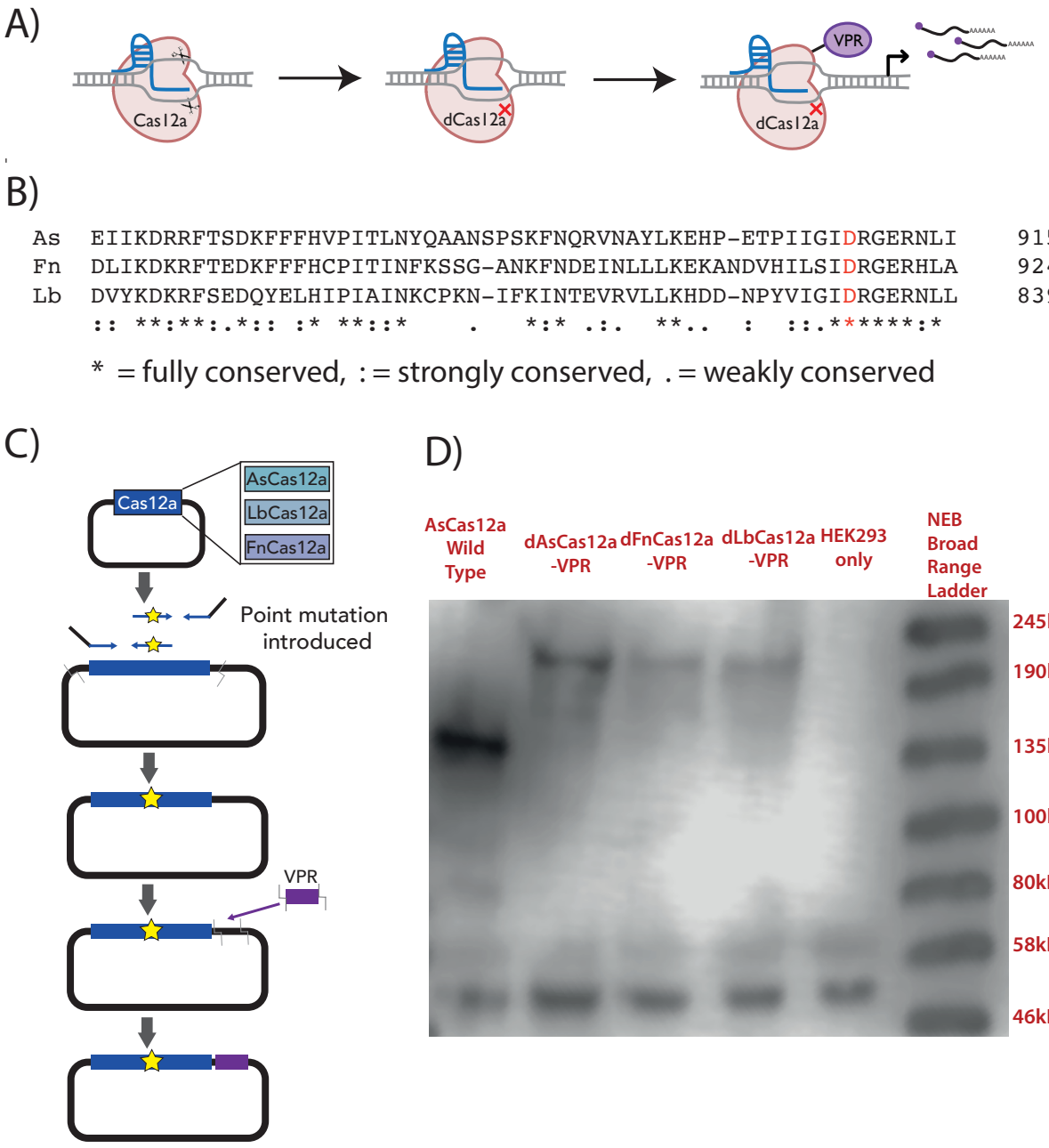

Figure 3-2 - Construction and expression of putative Cas12a-based synthetic transcription factors

A) Diagrammatic representation of the initial project objectives. B) Amino acid sequence alignment using Clustal Omega was performed for the three chosen variants of Cas12a; As, Fn and Lb. The conserved aspartic acid residue reported to abolish cleavage activity is highlighted in red and mutated to Alanine to abolish DNase activity. C) The cloning strategy for the generation of putative synthetic transcription factors. The amino acid substitution previously highlighted is introduced using multichange isothermal mutagenesis. After sequence verification the VPR transactivation domain was cloned to the 3'-end of the coding sequence, using restriction ligation. D) Western Blot analysis using anti-HA antibody showing the putative synthetic transcription factors tagged with HA at the expected sizes in HEK293 transformed cells. Wild-type AsCpf1 was used as a positive control and untransformed HEK293 cells (HEK293 only) served as negative control.

3.3 Generation of dCas12a-VPR C-terminal fusion proteins

The dCas12a plasmids were sequence verified before the transactivation domain VPR (Chavez *et al*., 2015) was sub-cloned to the 3'-end of the dCas12a coding sequence. The VPR domain was chosen, because it was demonstrated to be the most powerful single component transactivator domain in mammalian cells to date (Chavez *et al*., 2015). To generate this construct a traditional restriction ligation approach was used (Materials and Methods 2.1.9 and 2.1.2). First, the VPR domain was amplified with primers (VPR F + VPR R) (Table 2.1) designed to incorporate FseI and XhoI restriction sites, amplifying from the dCas9-VPR plasmid (Addgene #63798) (Table 2.2). The dCas12a plasmids and the amplicon were digested with FseI and XhoI. Subsequently the digested fragments were purified and ligated into the linearised dCas12a plasmids (Figure 3.2C). The resulting plasmids were sequence verified prior to protein expression analysis in transiently transfected HEK293 cells. The plasmids were transfected into HEK293 cells in a 24-well plate using Lipofectamine 2000 (ThermoFisher), with a non-transfected well serving as a mock transfection. The transiently transfected cells were incubated for 2 days at 37°C and 5% CO_2, before being washed with PBS and resuspended in SDS-PAGE sample buffer for Western Blot analysis (Materials and Methods 2.1.14) using mouse-derived anti-HA antibodies (Sigma HA-7; 019K4833). These antibodies bind to the retained HA tags within the dCas12a-VPR constructs (present in the original Cas12a plasmids) and enable visualisation of the HA-tagged proteins. The proteins could then be visualised after incubating the blot with anti-mouse HRP antibody. Specific bands of the predicted molecular weights of 213kDA, 213kDa and 205kDa for the As, Fn and Lb variants respectively were observed, along with the loading

control β-tubulin (51kDa) in all wells, as expected (Figure 3.2D). These results confirm that the full-length tagged proteins are successfully expressed in HEK293 cells. The fainter bands of the dCas12a-VPR variants compared to the wild-type AsCas12a band could be either owing to incomplete transfer of the fusion proteins from the SDS-PAGE gel to the membrane due to lower transfer speed seen for larger proteins, or due to the HA-tag being hidden in the middle of the dCas12a-VPR fusion variants, since the VPR domain was inserted to the 3'-end of the HA tag.

3.4 Screening functional activity of dCas12a-VPR C-terminal fusion proteins

Having successfully expressed the putative synthetic transcription factors for all three variants of dCas12a, we proceeded to test their functional activity in mammalian cells. The initial screen was designed to utilise a previously published Firefly luciferase-encoding reporter plasmid (FF Luc) (Kleinjan *et al.*, 2017), where 8 repeats of 20-nucleotide binding sequences were flanked by the SpCas9 PAM sequence on the 3'-end and the Cas12a PAM sequence on the 5'-end (Figure 3.3A). Upon sequence verification of the reporter construct, 2 of the 8 repeats were shown to contain mutations. However, we decided to proceed with the construct for screening, as 6 consecutive repeats should still provide sufficient recruitment of transactivation domains to observe transactivation activity. This enabled us to provisionally screen for functional activity of the constructs, using dCas9-VPR as a positive control.

First crRNAs were designed to enable targeting of the dCas12a-VPR variants to the Firefly Luciferase reporter plasmid. Due to the characterised PAM sequence of TTTN for the As and Lb variant and TTN for the Fn variant, the 'TTTT' PAM sequence was used 5' to the repeat sequence. The crRNAs for each variant were generated by ligating annealed oligos into the BpiI linearised U6 No scaf vector (As luc g F + As luc g R for the As crRNA, Fn luc g F + Fn luc g R for the Fn crRNA and Lb luc g F + Lb luc g R for the Lb crRNA) , which would enable expression of the incorporated crRNA sequences from the human U6 Pol III promoter (Figure 3.3B). The oligos were designed with a conserved 20 nucleotide spacer sequence, to target the repeat binding sites on the FF Luc plasmid, whilst incorporating the structural direct repeat unique to each variant being screened at the 5'-end. The direct repeat is unique for each variant and forms a hairpin which is recognised by the respective Cas12a variant. Finally, a single G was also included at the 5'-end of each oligo, corresponding to the start of the transcript, to improve expression of the crRNAs from the Pol III promoter (Gao et al., 2017).

The assay relied on the co-delivery of the reporter plasmid, alongside a control Renilla luciferase plasmid for normalisation of the activity. Each well in a 24-well plate of HEK293 cells was transiently transfected with the two luciferase plasmids alongside either a CRISPR derived transcription factor plasmid only (as a negative control), or the same plasmid and the respective targeting gRNA/crRNA plasmid (Figure 3.3B). We hypothesised that if the synthetic transcription factors were active, then the ratio of Firefly to Renilla luciferase would be considerably higher when the targeting gRNA/crRNA was present compared to when it was absent. The

Cas12a-derived putative synthetic transcription factors utilised the 'TTTT' PAM sequence, which fell within the initial characterisation by Zetsche *et al*. who described the PAM sequence as TTTN (Zetsche et al., 2015b). The transiently transfected HEK293 cells were incubated for two days at conditions previously described and washed with PBS, before addition of passive lysis buffer and incubation for 15 min at room temperature. The samples were then transferred to a 96-well plate for imaging with the Modulus II Microplate Multimode Reader (PROMEGA). The results observed for the positive control dCas9-VPR confirmed that the assay itself was working as expected, showing a significant increase in the ratio of Firefly to Renilla luciferase when the targeting gRNA was present. However, none of the dCas12a-VPR variants induced up-regulation (Figure 3.3C).

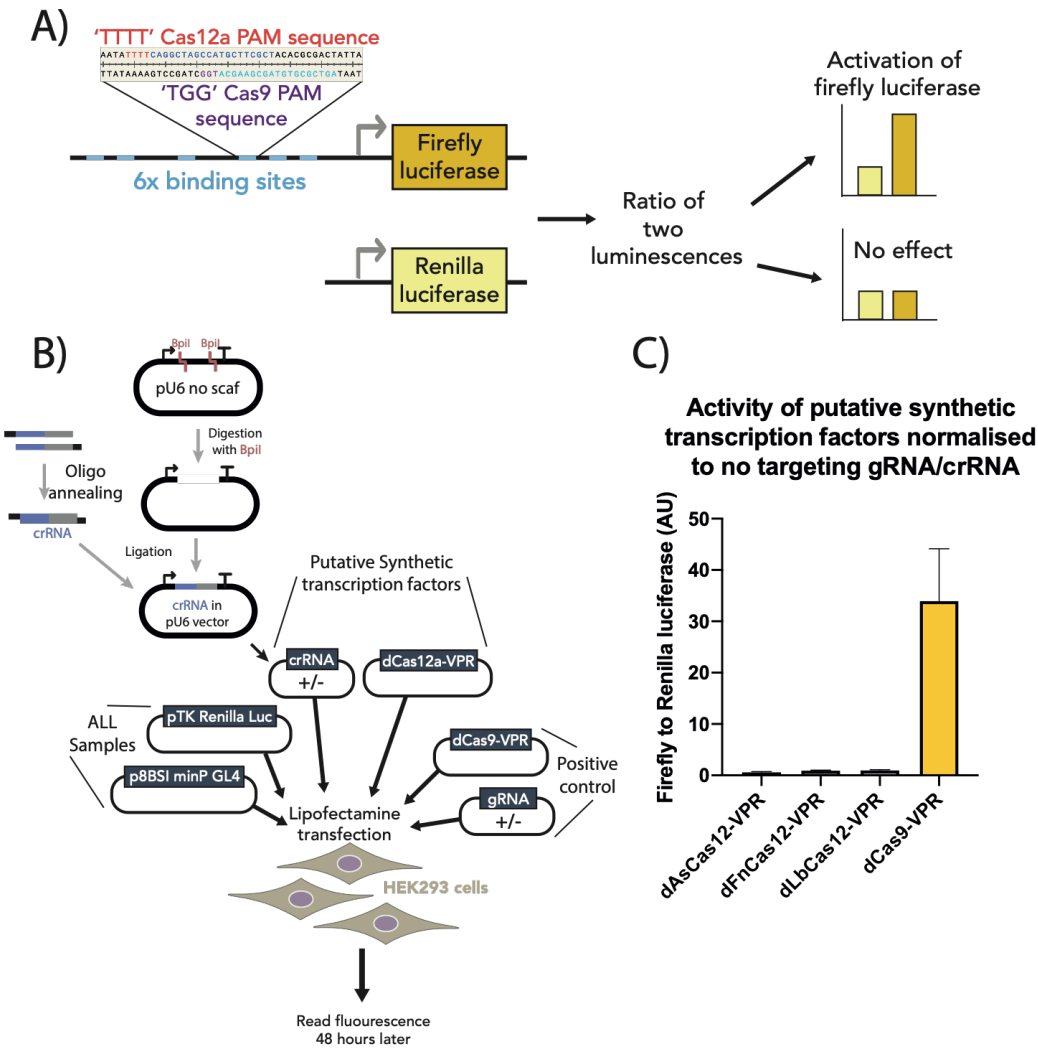

Figure 3-3 - Design and screening of putative dCas12a (c-terminally tagged) synthetic transcription factors

A) Diagrammatic representation of the reporter construct used in the dual luciferase assay, containing the repeated binding sites (blue) and the adjacent Cas12a (red) and Cas9 (purple) PAM sequences. The ratio of the two luciferase activities (expressed from separate plasmids) was read to assess the functional activity of the putative synthetic transcription factor in up-regulating the targeted Firefly Luciferase. B) Diagrammatic representation of the experimental setup to screen for activity of the putative synthetic transcription factors. The crRNA vectors were generated by linearising the U6 vector with BpiI before ligating the annealed oligos. All wells were transfected with the Luciferase constructs (pTK Renilla Luciferase and p8BSI minP GL4), the dCas12a-VPR constructs with or without their respective crRNAs, or dCas9-VPR with or without a targeting gRNA to serve as positive control. The Firefly and Renilla luciferase activities were measured 48 h post transfection. C) Graph showing the ratios of the Firefly to Renilla luciferase for the samples with the targeting crRNA/gRNA relative to the respective samples without a targeting crRNA/gRNA (n = 3) and error bars showing standard deviation.

3.5 Screening of unique targeting crRNAs

To understand the lack of discernible activity in the dCas12a-VPR variants, the first factor considered was the spacer sequence and the specific location of the crRNAs, with previous work showing that positioning in particular can have a large effect on activity (Gilbert et al., 2014). As previously mentioned, 2 of the 8 repeats of the targeted reporter construct were mutated in the dual luciferase assay, however they still preserved the 'TTTT' PAM sequence. Additionally, a third sequence provided a targeting sequence with a 'TTTT' PAM immediately downstream of the tandem repeats (Figure 3.4A). As such this presented 3 unique targeting sequences, allowing for a diversity of spacer sequences and positions to be screened. Screening of unique crRNAs focussed on the As or Lb variants of Cas12a-VPR, because they were characterised to have robust cleavage activity in mammalian cells (Zetsche et al., 2015b). The three crRNAs for both As and Lb were generated by restriction ligation of annealed oligos (Table 3.1) into a Bpi1 digested pU6 expression vector and subsequently sequence verified (Materials and Methods 2.1.1, 2.12 and 2.1.9). The dual luciferase assay was performed as previously described, however for this screen, the three newly constructed targeting crRNAs (each on separate plasmids) were co-transfected alongside their respective dCas12a-VPR variant. After 2 days, the cells were washed with PBS and the luciferase assay was performed (Materials and Methods 2.3.1). The results once again demonstrated that the assay was working as expected, with the dCas9-VPR showing robust up-regulation. However, the dCas12a-VPR variants showed no significant up-regulation (Figure 3.4B), suggesting it may not be the crRNA composition or target location within the promoter region that was responsible for inactivity.

Table 3.1 - 3 unique crRNAs for luciferase targeting

crRNA name	Forward oligo	Reverse oligo
As g1	As g1 F	As g1 R
As g2	As g2 F	As g2 R
As g3	As g3 F	As g3 R
Lb g1	Lb g1 F	Lb g1 R
Lb g2	Lb g2 F	Lb g2 R
Lb g3	Lb g3 F	Lb g3 R

(Refer to materials and methods section (Table 2.1) for sequence information.)

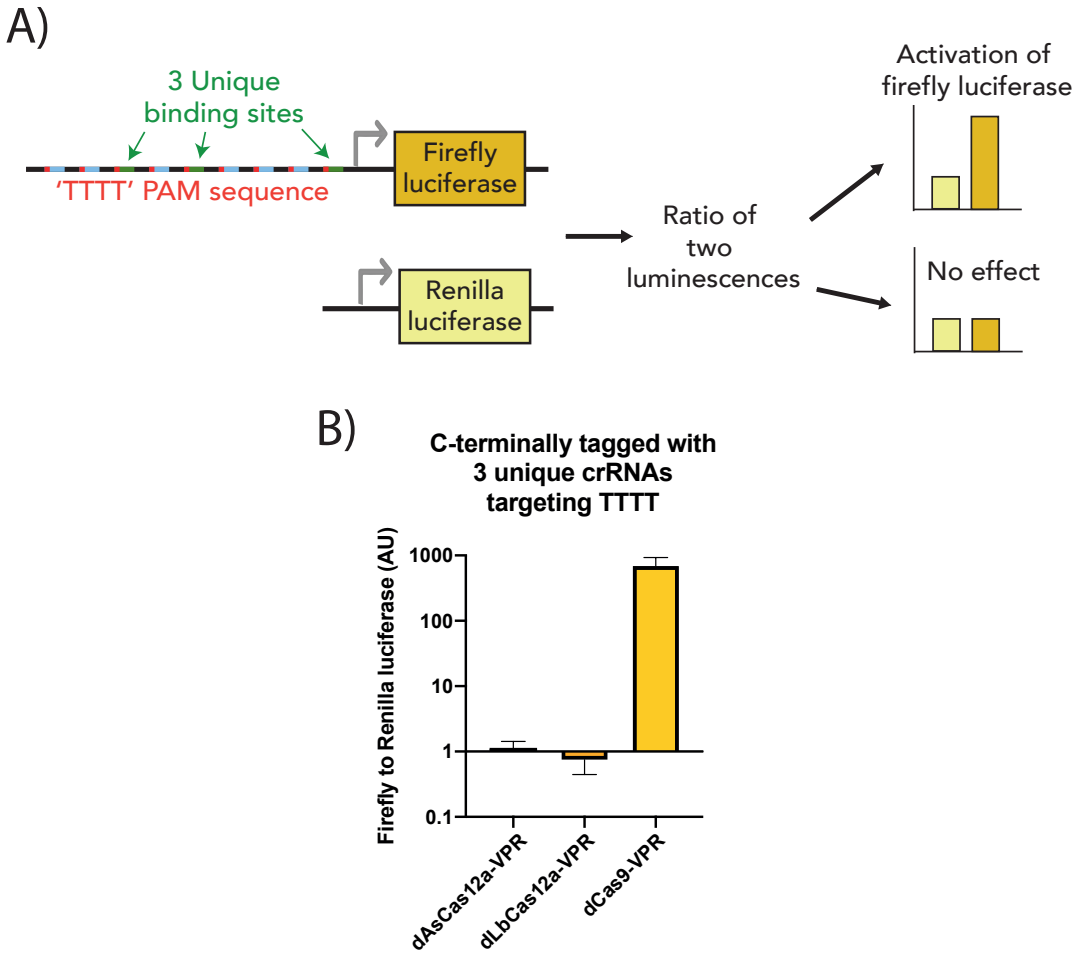

Figure 3-4 - Screening alternative targeting crRNAs

A) Diagrammatic representation of the unique locations being targeted. After three unique crRNAs were generated for the As and Lb Cas12a variants, the crRNAs were co-transfected alongside the two reporter plasmids and dAsCas12a-VPR or dLbCas12a-VPR. B) Results from dual luciferase assay, screening activity of dAsCas12a-VPR and dLbCas12a-VPR using the unique crRNAs, and dCas9-VPR with the original gRNA as a positive control (n = 3). All conditions are internally normalised to the respective constructs without the targeting crRNA/gRNA, with error bars showing standard deviation.

3.6 Generation of VPR-dCas12a N-terminal fusion proteins

Next, we investigated if the observed inactivity of dCas12a-VPR variants could be a result of the location of the effector protein tagged to the C-terminal of the dCas12a protein. We visualised the crystal structure of the active AsCas12a protein (Yamano *et al*., 2016) using PyMOL and found that the C-terminus and the N-terminus were exposed at opposite sides of the protein (Figure 3.5A). This suggested that any detrimental effect of C-terminal fusion of Cas12a with the VPR domain through misfolding can potentially be ameliorated by N-terminal fusion. This strategy was supported by the success with dCas9, where tagging VP64 domain to the N-terminus preserved the ability to activate target genes (Duellman *et al*., 2017). To test this hypothesis, we N-terminally tagged dAsCas12a using Gibson assembly (Materials and Methods 2.1.13). The 5'-end of the dAsCas12a coding sequence was excised from the dAsCas12a plasmid by digesting with KpnI and SbfI, generating a linearized plasmid backbone. Then, the VPR domain was amplified from dCas9-VPR (F primer VPR + R VPR AsCpf1) and the 5'-end of the coding sequence for dAsCas12a was amplified from the dAsCas12a plasmid (N term As F + N term As R) (Figure 3.5B). The VPR domain was amplified using primers designed to incorporate the flexible glycine serine rich linker on the 3'-end of the VPR domain. Furthermore, the primers also incorporated 40bp of homology to the linearized plasmid and the dAsCas12a-derived amplicon. Gibson assembly was performed to generate the N-terminally VPR tagged variant of dAsCas12a, with a flexible linker between the two proteins (Figure 3.5B). The assembled construct was sequence verified prior to functional activity screening.

3.7 Screening of N- and C-terminally VPR tagged dAsCas12a fusion proteins

The N-terminally tagged VPR-dAsCas12a was transfected alongside the C-terminally tagged dAsCas12a-VPR variant, each co-transfected alongside the first targeting crRNA tested (Results 3.4). The results were normalised to delivery without a targeting gRNA/crRNA. After 2 days, the cells were washed with PBS and the luciferase assay was performed. The dual luciferase assay demonstrated that whilst the dCas9-VPR positive control was able to up-regulate Firefly luciferase relative to Renilla luciferase, neither the C-terminal nor the N-terminal tagged variant of dAsCas12a-VPR fusion proteins showed activation (Figure 3.5C).

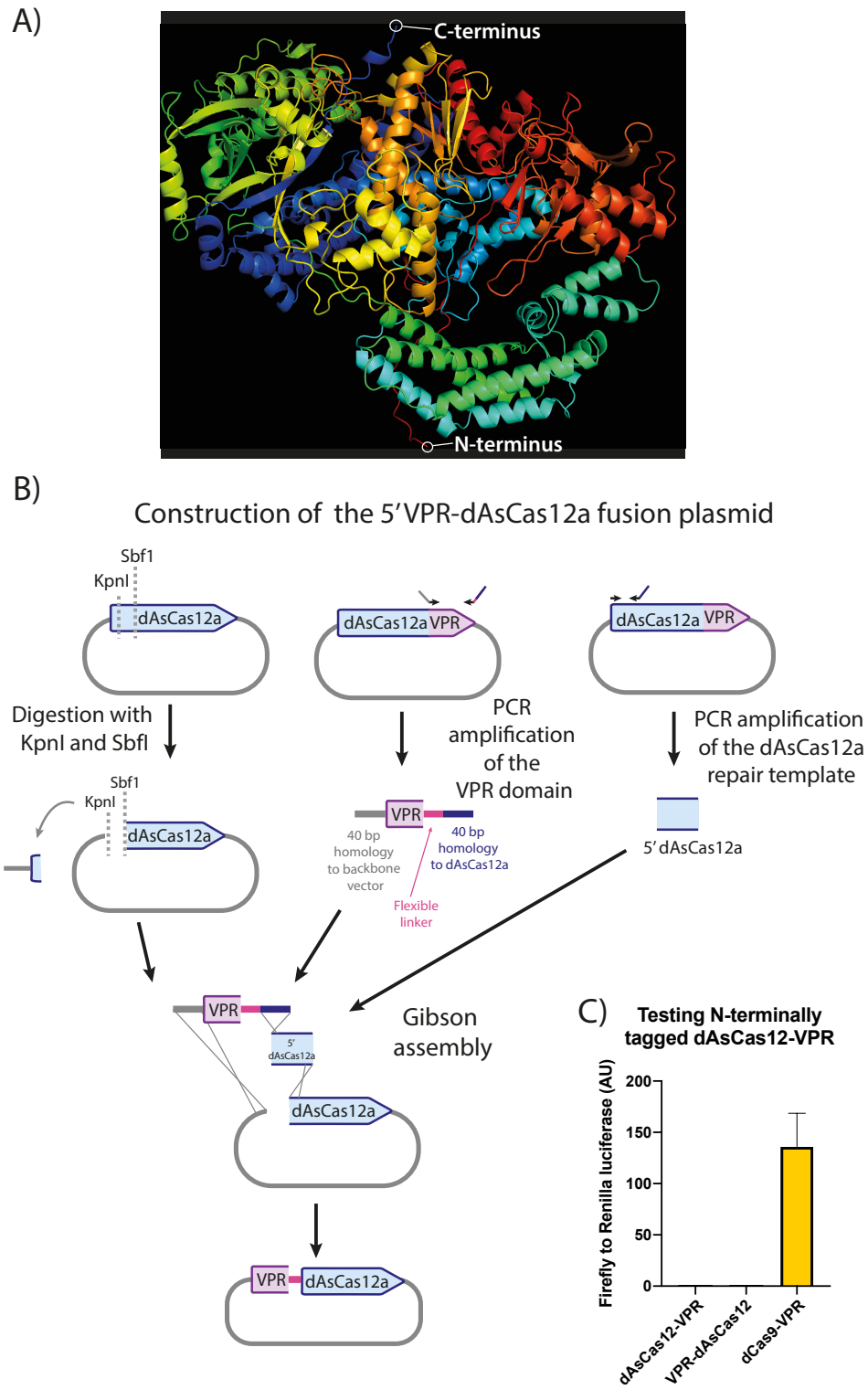

Figure 3-5 - Design and testing of N-terminally tagged dAsCas12a-based synthetic transcription factors (VPR)

A) Schematic representation of wild-type AsCas12a, with the C-terminus and N-terminus of the protein highlighted. B) Diagrammatic representation of the cloning strategy to generate the N-terminally tagged VPR-dAsCas12a. C) Results of dual luciferase assay screening in HEK293 cells for activity of the N-terminally tagged VPR-dCas12a, alongside the C-terminally tagged dCas12a-VPR and dCas9-VPR serving as a positive control (n = 3). Error bars display the standard deviation.

3.8 Generation of internally tagged dAsCas12a-VP64

As a final strategy, protein engineering involving the insertion of shorter VP64 domains into exposed loops of dAsCas12 was considered, with the shorter VP64 domain chosen to minimise potential disruption caused by insertion. This strategy was based on a report were oestrogen receptor domains were inserted at several locations of the Cas9 protein, enabling oestrogen-dependent Cas9 cleavage that is dependent on conformational changes in the oestrogen receptor domain (Oakes *et al*., 2016). They specifically noted that the successful insertions were disproportionately found around flexible loops. These findings suggest that it may be possible to insert domains within a related CRISPR homing endonuclease, without compromising the activity of either the homing endonuclease or the inserted domains. As such, we explored the protein structure of AsCas12a and identified two candidate loci with exposed flexible loop regions (Figure 3.6A). Instead of a full length VPR domain, this strategy centred upon inserting a shorter VP64 domain, which would typically have reduced activity, but could also decrease the likelihood of compromising the folding and structure of the dAsCas12a protein it would be inserted into. These constructs were generated using Gibson assembly (Materials and Methods 2.1.13), first digesting and removing the 3'-end of the dAsCas12a coding sequence, before reconstituting the coding sequence, using three amplicons with overlapping overhangs (Figure 6B). Two of these three amplicons were derived from the N-terminal portion (As PmlI F + As Site 1 PmlI R for site 1 and As PmlI F + As Site 2 PmlI R for site 2) and C-terminal portion (As Site 1

BamHI F + As BamHI R for site 1 and As Site 2 BamHI F + As BamHI R for site 2) of the original dAsCas12a vector. The final amplicon corresponding to the VP64 domain was amplified from the dCas9-VPR vector (VP64 Site1 5AA F + VP64 Site1 5AA R for site 1 and VP64 Site2 5AA F + VP64 Site2 5AA R for site 2). The primers used for amplification of the VP64 domain incorporated flexible 5AA linkers and 40bp of homology to the dAsCas12a amplicons (Figure 3.6B).

3.9 Screening of internally tagged dAsCas12a fusion proteins

After sequence verification, the internally tagged dAsCas12a-VP64 variants were transfected into HEK293 cells alongside the dCas9-VPR as positive control. The constructs delivered without a targeting gRNA/crRNA served as negative controls. After 2 days, the cells were washed with PBS and the luciferase assay was performed. The results showed no significant upregulation for either of the two dAsCas12a variants with VP64 insertions at different sites, but significant up-regulation for the dCas9-VPR positive control (Figure 3.6C).

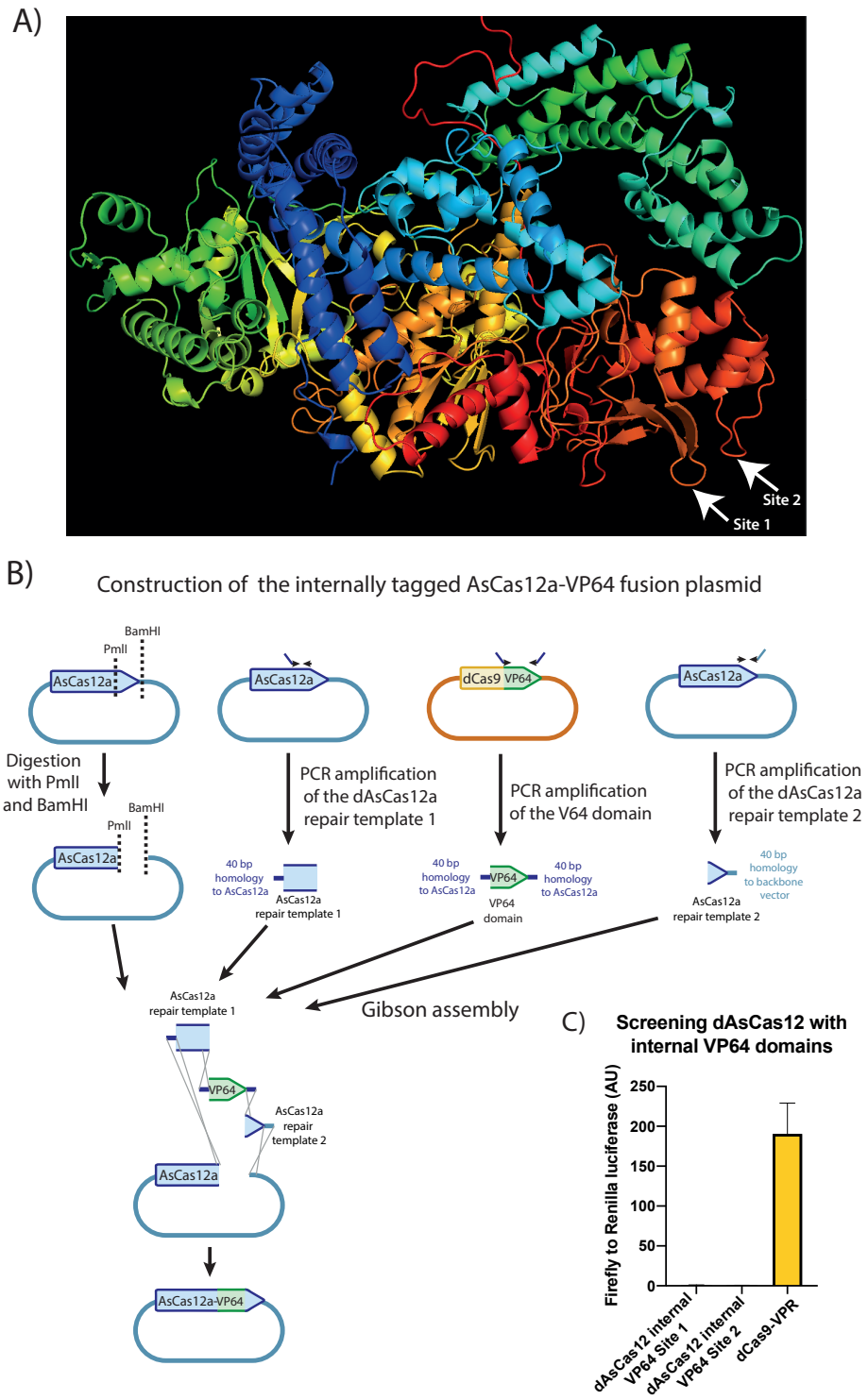

Figure 3-6 - Design and screening of internal [VP64] tagged dAsCas12a-based synthetic transcription factors

A) Schematic representation of wild-type AsCas12a, with the flexible loops targeted for VP64 insertion at site 1 and site 2 highlighted. B) Diagrammatic representation of the cloning strategy to generate the two internally tagged dAsCas12a-VP64 variants. C) Results of dual luciferase assay screening for Site 1 or Site 2 internally tagged dAsCas12a-VP64 variants, with dCas9-VPR serving as positive control (n = 3). Error bars display the standard deviation.

3.10 Screening an alternate internally tagged fusion proteins

Whilst these insertion sites had not resulted in active dCas12a-derived synthetic transcription factors, findings from a collaborator suggested that an alternative insertion site could enable generation of an active synthetic transcription factor (Dominguez-Monedero and Davies, 2018). Similar to the work by Oakes *et al*. (Oakes et al., 2016), Dominguez-Monedero and colleagues identified tamoxifen-inducible activity of Cas12a variants when an oestrogen receptor was inserted within AsCas12a. We generated a new plasmid with VP64 inserted, with no flanking linkers, into the exposed flexible loop identified by Dominguez-Monedero *et al*. (Figure 3.7A), using the same Gibson assembly strategy previously described (Figure 6B). As before, the N-terminal fragment of dAsCas12a was amplified (Outer Alaz As F + Inner Alaz As R), as was the C-terminal fragment (Inner Alaz As F + Outer Alaz As R). VP64 was amplified (VP64 Alaz As F + VP64 Alaz As R) with primers designed to incorporate 20bp overhangs with the N-terminal and C-terminal dAsCas12a amplicons. dAsCas12a was digested with SbfI and PmlI and run on a gel. The linearised plasmid backbone gel extracted (Materials and Methods 2.1.11) and Gibson assembly (Materials and Methods 2.1.13) was performed with the three purified amplicons.

After sequence verification, this construct was transfected alongside the N-terminally tagged dAsCas12a-VPR and dCas9-VPR serving as a positive control into HEK293 cells. After 2 days, the cells were washed with PBS and the luciferase assay

was performed. The dual luciferase assay was performed as previously described. However, we observed no transactivation with the dAsCas12a constructs (Figure 3.7B).

Taken together these results strongly suggest that the point of failure for these putative synthetic transcription factors was not likely the positioning or composition of the crRNAs, or the positioning of effector domains relative to dCas12a.

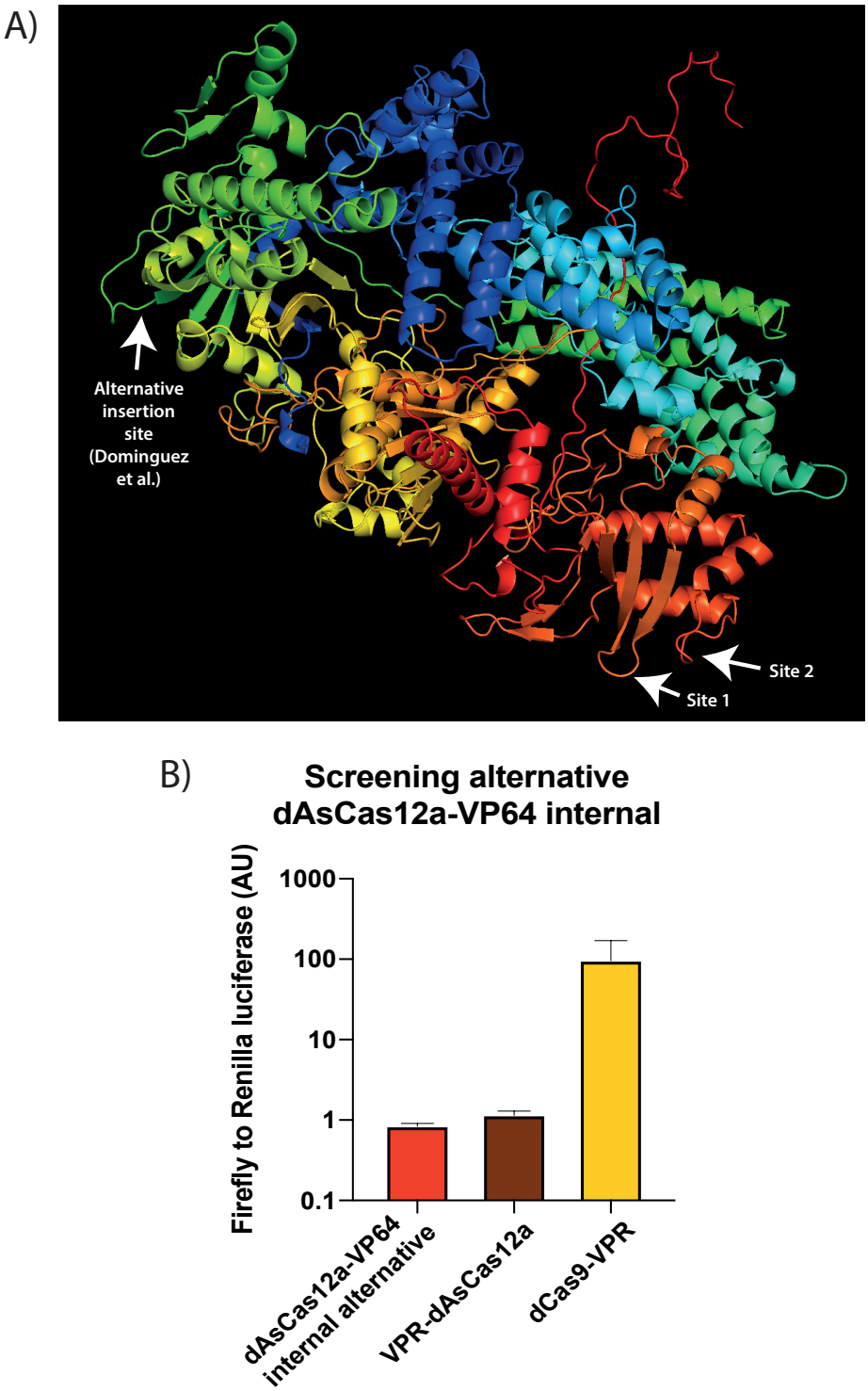

Figure 3-7 - Design and screening of an alternative variant of internal [VP64] tagged dAsCas12a-based synthetic transcription factors (based on collaboration work with Davies group)

A) Schematic representation of wild-type AsCas12a with the original flexible loops targeted for VP64 insertion at site 1 and site 2 highlighted, alongside the alternative site recommended by a collaborator. B) Results of dual luciferase assay screening for dAsCas12a-VP64 variant tagged at an alternative site, with N-terminally tagged VPR-dAsCas12a serving as a negative control and dCas9-VPR serving as a positive control (n = 3). Error bars display the standard deviation.

3.11 Modification of PAM sequence and screening with dAsCas12a variants

With none of the protein engineering strategies proving successful, we focussed on dissecting the importance of the targeting sequence and more specifically the chosen PAM sequence. As previously highlighted, Zetsche *et al*. had characterised the PAM sequence as TTTN for As and Lb variants and TTN for the Fn variant (Zetsche et al., 2015a). However, subsequent publications suggested that this may represent a subtle mischaracterisation, with Leenay *et al*. performing a comprehensive screen with the DNase inactive variant of FnCas12a that showed the PAM sequence for Fn was better represented by the motif 'TTV' (with V standing for any nucleotide, but T) (Leenay et al., 2016). This preference of V for the 3'-end nucleotide within the PAM sequence was also shown to translate to AsCas12a and LbCas12a by Kim *et al*. who observed a preferred PAM sequence of 'TTTV' for both variants (Kim *et al*., 2017). This offered a compelling hypothesis explaining why none of the previously described protein engineering strategies were successful, as the dual luciferase screen had relied upon a PAM sequence of 'TTTT'. As such a new crRNA was designed and constructed for screening the C-terminally and N-terminally VPR tagged dAsCas12a variants. The target sequence was moved by a single nucleotide at the 3'-end on the targeted reporter plasmid, to provide a PAM sequence of 'TTTC' (Figure 3.8A). The corresponding crRNA vector was generated by annealing crRNA oligos (Alt PAM As Luc F + Alt PAM As Luc R) (Table 2.1) and ligating into a gel extracted BpiI digested U6 no scaf vector (Materials and Methods 2.1.1, 2.1.2 and 2.1.11).

Using this new crRNA for the As variants, we re-screened both the C-terminally and N-terminally tagged dAsCas12a-VPR using the dual luciferase assay. We were now able to see robust up-regulation for both constructs, with comparable up-regulation to dCas9-VPR, with a mean of 48- and 128-fold increase for the C-terminally and N-terminally tagged variants, respectively, compared to a mean 89-fold increase for the positive control dCas9-VPR (Figure 3.8B). When a Welch's t-test was performed, no significant difference in activity between the C-terminally and N-terminally tagged dAsCas12a-VPR was observed ($P = 0.28$).

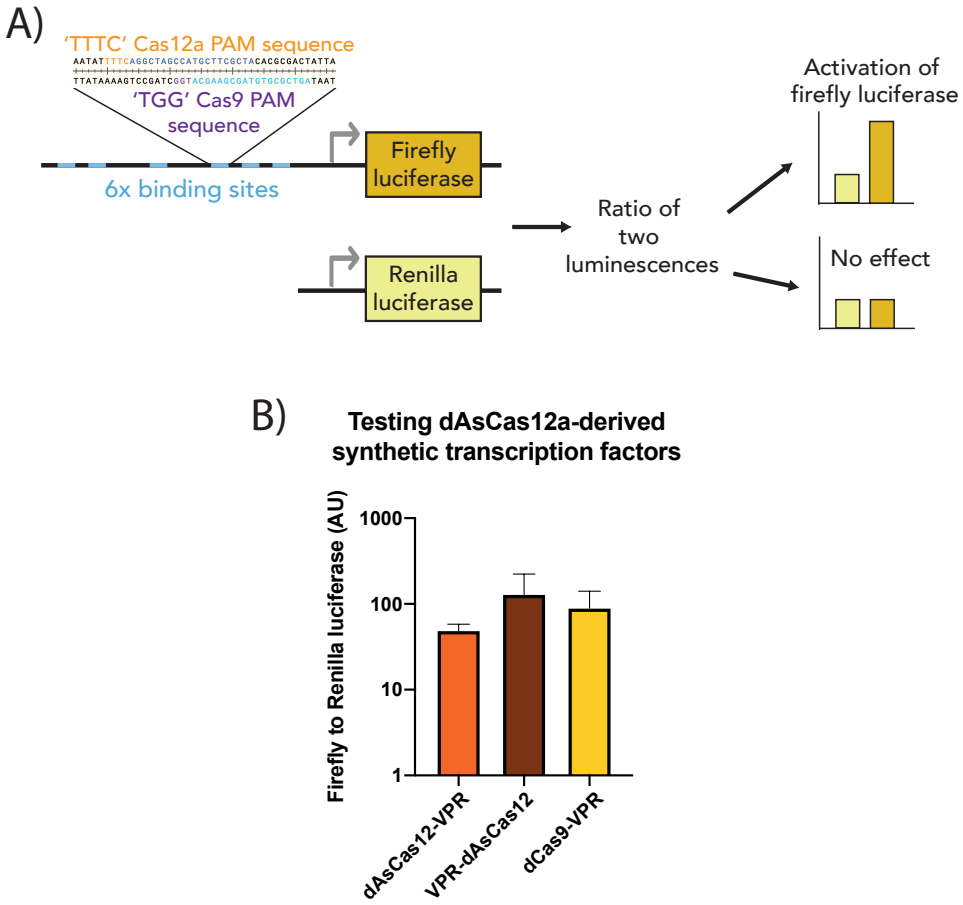

Figure 3-8 - Screening N-terminally tagged dAsCas12a VPR against TTTC PAM

A) Diagrammatic representation of the dual luciferase assay, highlighting the modified target sequence utilising the 'TTTC' PAM sequence. B) Results of dual luciferase assay screening activity for dAsCas12a- variants C-terminally and N-terminally tagged with VPR and targeted with the alternative crRNA, with dCas9-VPR serving as a positive control (n = 3). Error bars display the standard deviation.

3.12 Conclusion and discussion

A number of protein strategies were attempted for the generation of Cas12a derived synthetic transcription factors, however the PAM sequence appeared to be the critical bottleneck in development. These challenges highlight the importance improved representations of the binding/cleavage capabilities of CRISPR effectors (Leenay et al., 2016). Such approaches help to ameliorate the challenges that emerge when describing a PAM sequence using letters, which invariably lead to a conflict between flexibility (describing non-canonical PAM sequences) and stringency (avoiding inclusion of PAM sequences that appear to show no significant binding/cleavage).

In the preceding work we used a dual luciferase assay to show successful generation of two different dAsCas12a-derived synthetic transcription factors employing two unique strategies for generation (C-terminal and N-terminal tagging). This shows that dCas12a is able to tolerate the fusion of effector domains to the N-terminus as well as the C-terminus of the protein, similar to what has been observed for dCas9 dCas9 (Duellman et al., 2017). We have also been able to confirm, what was subsequently published, that AsCas12a disfavours a T at the 3'-position of the PAM sequence in mammalian cells, but permits a target sequence of TTTC when targeting a plasmid within HEK293 cells (Kim et al., 2017). Moving forward, we will look into further characterising the generated synthetic transcription factors, giving particular attention to the remaining two Cas12a variants (Fn and Lb) and screening for activity when targeting endogenous promoters.

Chapter 4 Expanding and characterising the Cas12a tool

4.1 Introduction

Work published by Tak et al. has shown the capacity of both dAsCas12a-VPR and dLbCas12a-VPR to enable transactivation of target endogenous genes in mammalian cells (Tak et al., 2017). They also demonstrate significantly higher transactivation can be achieved by delivering pools of multiple crRNAs targeting a single gene. However, their work only focusses on two Cas12a variants and also does not seek to address the question of the level of orthogonality observed between these two different variants. To the best of our knowledge, the question of orthogonality between different Cas12a variants has yet to be explored in the literature and no published work has shown that FnCas12a can be converted into a synthetic transcription factor. In the following chapter we aimed to test whether all three variants of Cas12a tested (As, Fn and Lb) could be converted into synthetic transcription factors and enable transactivation in a mammalian context when a modified 'TTTV' PAM sequence was targeted. As orthogonality has been demonstrated between Cas9-derived from different species (Esvelt et al., 2013) (Gao et al., 2016), it is reasonable to hypothesise that Cas12a variants from different species may also operate orthogonally. The crRNA that targets the Cas12a protein is composed of a direct repeat sequence upstream of a spacer sequence (Figure 4.1a). The direct repeat sequence forms a stem loop and interacts with the Cas12a protein and the spacer sequence confers targeting specificity and can hybridise with the associated DNA sequence. If two Cas12a variants were orthogonal then the direct repeat derived from one of the species would only interact and enable targeting for the native protein variant, not the second variant and this would also be the case for the second variant (Figure 4.1b).

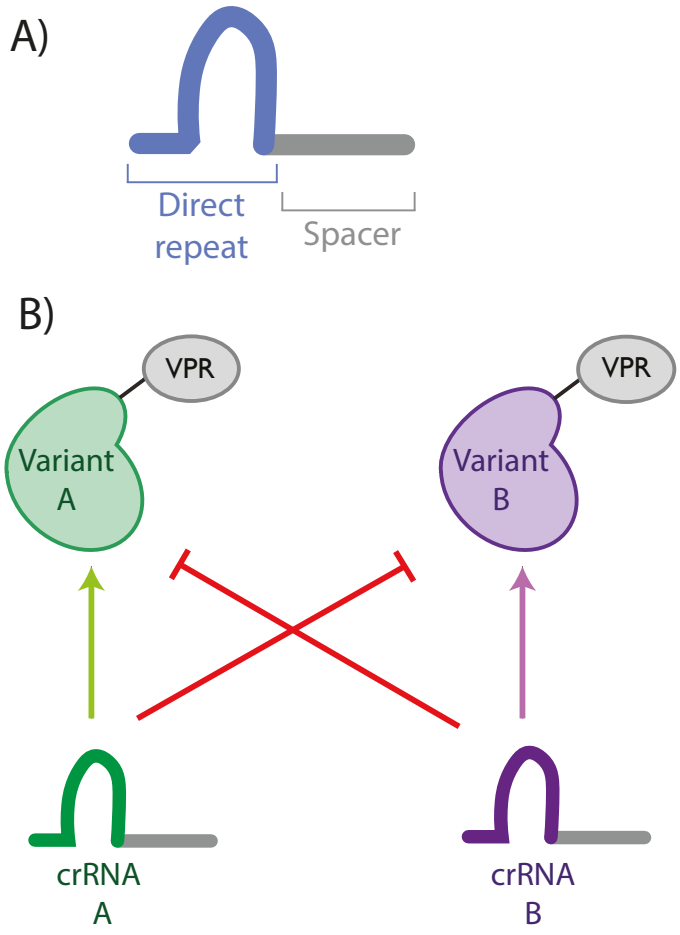

Figure 4-1 - Screening of three variants of dCas12a-VPR synthetic transcription factors against TTTC PAM

A) Diagram showing the composition of the crRNA. The crRNA is composed of a structural direct repeat at the 5' end, which interacts with the Cas12a and the spacer sequence encoding the targeting sequence. B) Diagram representing orthogonality between two different dCas12a-VPR variants. For an orthogonal pair, each crRNA will only interact with it's respective dCas12a-VPR variant.

Due to the applications dependent upon orthogonality (described in Chapter 1) we also screened for the orthogonality of the different dCas12a-VPR variants, specifically testing whether the three different variants demonstrated significant activity when targeted with a crRNA derived from one of the remaining two species (e.g dAsCas12a-VPR targeted by an Fn or Lb crRNA). Finally we also aimed to test if some of the effects observed for As and Lb variants (transactivation of endogenous genes and improved activity when multiple crRNAs were delivered) were translatable to the Fn variant.

4.2 Screening dFnCas12a-VPR and dLbCas12a-VPR against a TTTC PAM

Having demonstrated in chapter 3 that dAsCas12a-VPR was able to transactivate a Firefly luciferase when targeting the PAM sequence 'TTTC', we next determined whether dFnCas12a-VPR and dLbCas12a-VPR were also able to transactivate the targeted Firefly luciferase when a 'TTTC' PAM sequence was targeted (Figure 4.2A). To test this a new crRNA was designed and generated for each dCas12a-VPR variant, utilising the same spacer sequence previously used for screening dAsCas12a-VPR (described in chapter 3). These crRNAs were generated from annealed oligos containing the respective direct repeat and spacer sequences (Alt PAM Fn Luc F + Alt PAM Fn Luc R and Alt PAM Lb Luc F + Alt PAM Lb Luc R) (Table 2.1), before ligating these into a Bpi1 digested pU6 no scaf plasmid, to enable expression of the crRNAs. After sequence verification of these alternative crRNAs, the plasmids were transfected into HEK293 cells with their respective dCas12a-VPR plasmids, and the Firefly and Renilla Luciferase reporter plasmids for activity screening. These two conditions were screened alongside dAsCas12a-VPR (C-terminally tagged) and dCas9-VPR, serving as positive controls, with all conditions being compared to the respective synthetic transcription factors without a targeting gRNA/crRNA as an internal negative control. As previously described, the dual luciferase assay was performed 2 days post transfection. The results confirmed that dAsCas12a-VPR was able to robustly transactivate the targeted Firefly Luciferase (Figure 4.2B), with a Welch's t-test showing significant transactivation had been achieved ($P = 0.0013$). Furthermore, we also observed significant transactivation for the Fn ($P = 0.02$) and Lb ($P = 0.04$) variants alongside the dCas9-VPR positive control. These results were of particular interest, as previous studies had suggested that FnCas12a was unable to cleave DNA in mammalian cells (a finding later refuted (Tóth et al., 2018)). Furthermore, the Fn variant has been characterised to have a shorter PAM sequence of

'TTV' rather than 'TTTV' for the As and Lb variants, meaning this variant can target genomic loci more densely. Denser targeting aids a variety of applications, where there are restrictions on the number of available targeting sites (such as when the targeting window is constrained or multiple promoters all need to be targeted) or where the targeting of more sites is desirable (if multiple crRNAs show higher transactivation or synergy).

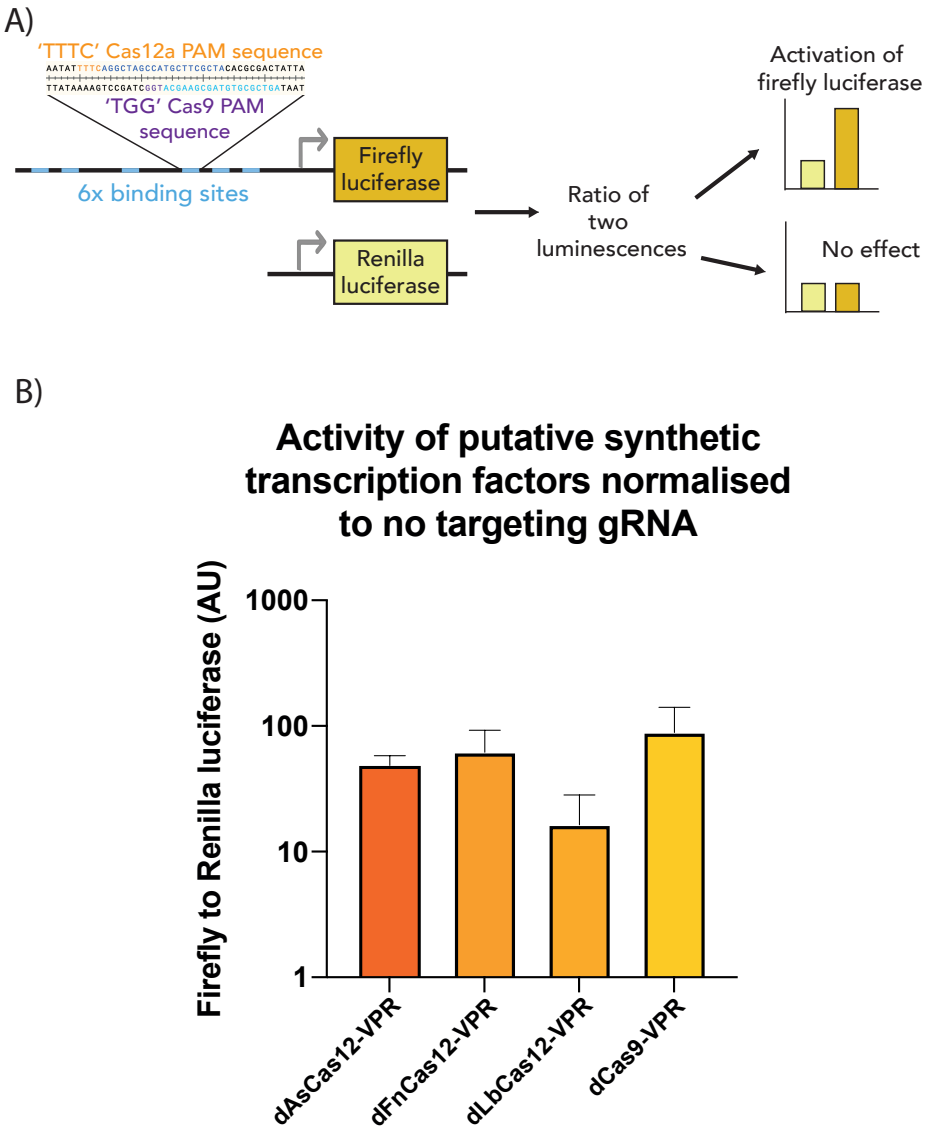

Figure 4-2 - Screening of three variants of dCas12a-VPR synthetic transcription factors against TTTC PAM

A) Diagrammatic representation of the dual luciferase assay. The repeated binding sites are shown in blue, with the targeted 'TTTC' PAM sequence shown in orange. B) Graph showing the ratios of the Firefly to Renilla luciferase for the samples with the targeting crRNA/gRNA relative to the respective samples without a targeting crRNA/gRNA (n = 3). Error bars display the standard deviation.

4.3 Screening for orthogonality between dCas12a-VPR variants

Having demonstrated dCas12a-VPR derived from three different species were able to induce robust transactivation, we next explored the orthogonality and fidelity of each variant relative to one another. The question of orthogonality specifically seeks to determine whether two variants will show activity with a crRNA specific to its own species, but not a crRNA derived from the other species. For example, if dAsCas12a-VPR only showed activity when delivered with a crRNA possessing the As derived direct repeat, not with a crRNA possessing the Fn derived direct repeat and dFnCas12a-VPR only showed activity when delivered with a crRNA possessing the Fn derived direct repeat, these would be described as orthogonal. The question of fidelity seeks to explore whether an individual variant only displays significant activity when delivered with a crRNA derived from its own species and not from crRNA derived from different species.

As the direct repeat sequence for all three crRNAs was highly conserved, with a difference of only 1-3 bp localised to the hairpin loop (Figure 4.3A), we hypothesised that there may be some cross-reactivity between the different dCas12a-VPR/crRNA pairs. We screened the ability of each dCas12a-VPR variant to transactivate a targeted reporter gene when co-transfected with crRNAs with a direct repeat derived from another species compared to one derived from its own species (Figure 4.3B).

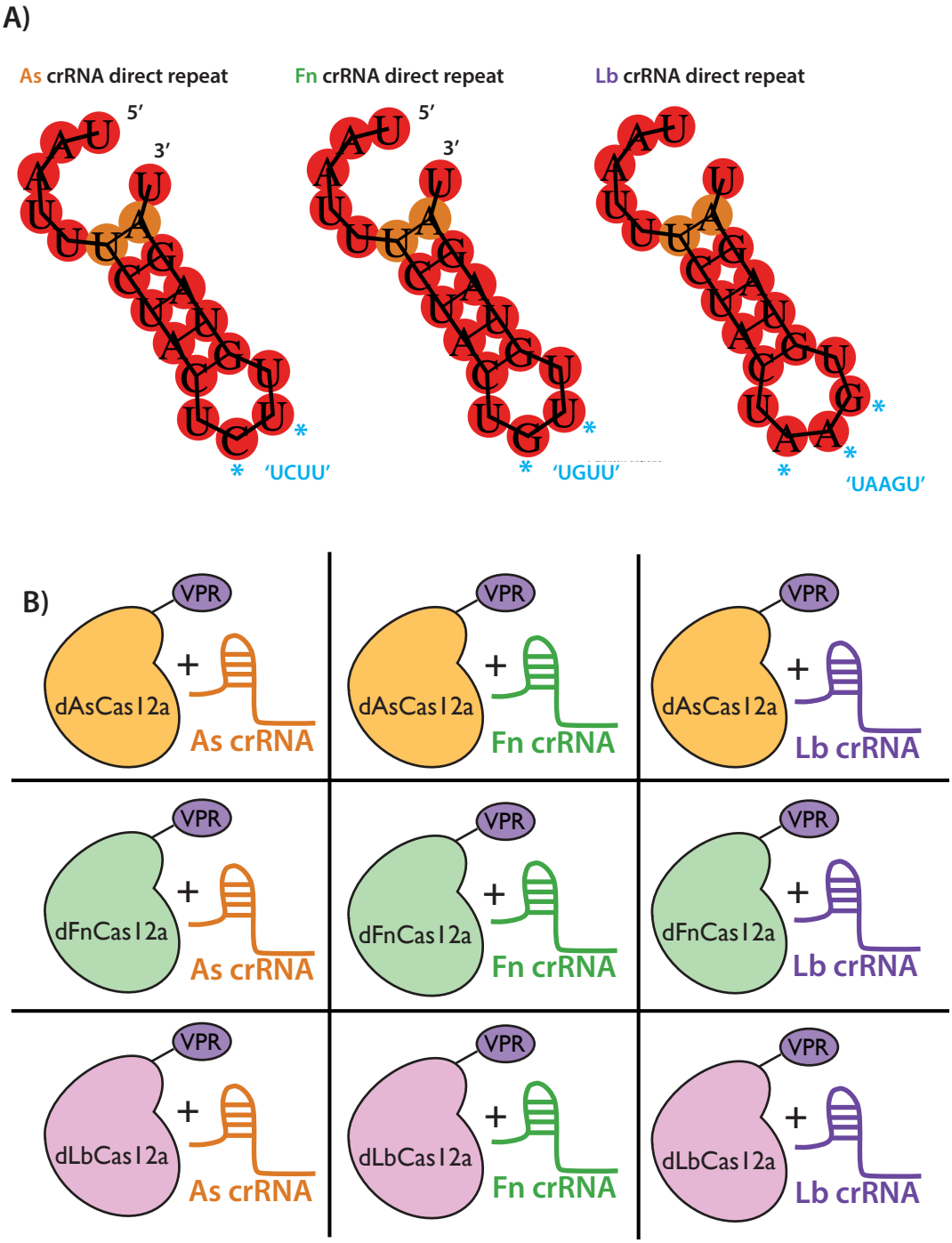

Figure 4-3 - Design for orthogonality screen of dCas12a-VPR variants

A) Minimal free energy structure of direct repeat sequences derived from As, Fn and Lb species predicted using RNA fold. The 5' to 3' sequence composition of the hairpin loop is shown beneath the predicted structure and unique nucleotides are highlighted with blue asterisk B) Diagrammatic representation of the combinations of dCas12a-VPR and targeting crRNAs to be screened using the dual luciferase assay, targeting a 'TTTC' PAM sequence.

To test for orthogonality, each dCas12a-VPR variant was transfected with the Firefly and Renilla Luciferase reporter plasmids into HEK293 cells alongside each crRNA to test all possible dCas12a-VPR/crRNA combinations (Figure 4.3B). Each dCas12a-VPR variant was also transfected without a targeting crRNA to serve as a negative control. Two days after transfection the dual luciferase assay was performed. For the purposes of screening we used an ordinary one-way ANOVA, setting significance at P = 0.05. When performing a Tukey's multiple comparison test we observed significant transactivation over the negative control (no crRNA) for dAsCas12a-VPR and dFnCas12a-VPR when targeted with their own crRNAs (P = 0.0003 and P = 0.006 respectively) (Figure 4.4A). However, for dLbCas12a-VPR the P value 0.051 lay just below the threshold of significance. We were able to see that dAsCas12a-VPR was also able to induce significant transactivation when targeted by the Fn crRNA (P = 0.005), which may in part be explained by the minimal difference in sequence composition for the As and Fn crRNAs direct repeat (Figure 4.3A). We could see high fidelity for the dFnCas12a-VPR variant, with a significantly higher transactivation seen when the construct was delivered with the Fn crRNA compared to the As crRNA (P = 0.01) or the Fn crRNA compared to the Lb crRNA (P = 0.008). We also observed that when assessing dLbCas12a-VPR delivered with Lb crRNA compared to As or Fn crRNA, we were able to see close to significantly higher transactivation for the Lb crRNA condition (0.054 and 0.052 respectively). The previously described data is visually summarised with the heatmap in figure 4.4B.

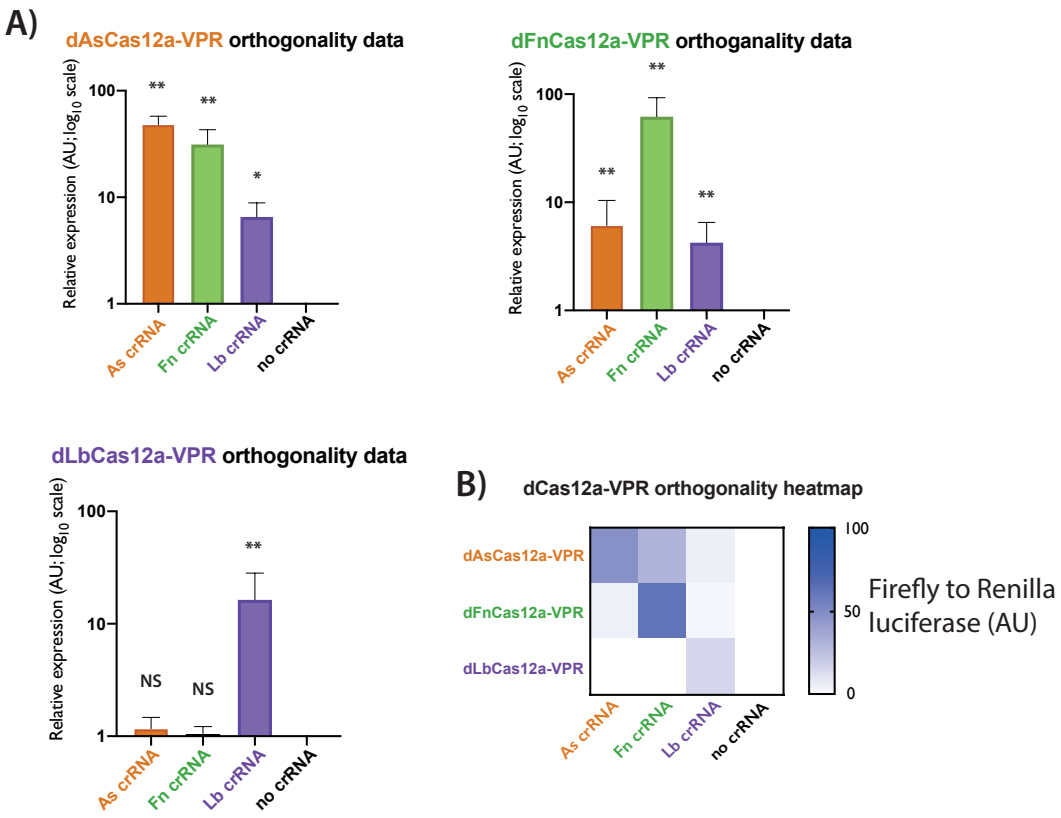

Figure 4-4 - Results for orthogonality screen of dCas12a-VPR variants

A) Graphs showing the relative expression of each of the three dCas12a-VPR variants screened for activity using the dual luciferase assay (n = 3). Each dCas12a-VPR variant was screened with crRNAs utilising a direct repeat derived from each of the respective species alongside screening with no crRNA as a negative control. Relative expression for each individual dCas12a-VPR variant is normalised to the no crRNA condition set to a value of 1. Error bars display the standard deviation. B) Heatmap representing the activity of each of the three different synthetic transcription factors paired with each of the three crRNAs as well as no crRNA. The darker squares represent higher relative activity.

4.4 Targeting dFnCas12a-VPR to endogenous mammalian promoters

Having demonstrated that dFnCas12a-VPR was able to transactivate a plasmid-based reporter, we next explored its ability to transactivate endogenous genes. At this time Tak et al. had published work showing that robust transactivation could be achieved for three different genes; *AR*, *HBB* and *NPY1RN* when targeting dLbCas12a-VPR to their promoters. We decided to utilise the same targeting crRNA arrays screened by Tak et al. (Tak et al., 2017), but substituting the Lb derived direct repeats for Fn derived direct repeats, to enable screening of activity for dFnCas12a-VPR (Figure 4.5A). Oligo pairs were designed for constructing each array (Table 4.1), incorporating an 8bp overhang between each oligo pair to facilitate ligation. Each of the three crRNA arrays were constructed and sequence verified before being subsequently transfected into HEK293 cells alongside dFnCas12a-VPR, with dFnCas12a-VPR only serving as a negative control for each target gene. Three days post transfection, total RNA was extracted from each sample and 1µg was used to synthesize cDNA using oligo dT(20) primers (further described in Materials and Methods – 2.3.2). 1µl of the generated cDNA was used for qRT-PCR analysis of the target gene and normalised to reference gene GAPDH (Figure 4.5B). The results were then analysed, performing a Welch's t test to compare the $\Delta\Delta$CT values between the array conditions and the negative control (dFnCas12a-VPR only) (Materials and Methods 2.3.2). We observed highly significant transactivation was achieved for all three genes; P = 0.001 for AR, P = 0.004 for HBB and P = 0.003 for NPY1R. Of note, whilst significant transactivation was achieved for all three arrays, the average fold up-regulation was substantially higher for HBB (~7745 fold increase) compared to AR (~2.8 fold increase) and NPY1R (~10 fold increase).

Table 4.1 - Fn adapted Tak arrays

AR array	HBB array	NPY1RN array
First AR MST F	HBB MST First F	NPY1R MST First F
First AR MST R	HBB MST First R	NPY1R MST First R
AR MST 2 F	HBB MST 2 F	NPY1R MST 2 F
AR MST 2 R	HBB MST 2 R	NPY1R MST 2 R
AR MST 3 F	HBB MST 3 F	NPY1R MST 3 F
AR MST 3 R	HBB MST 3 R	NPY1R MST 3 R
AR MST Last F	HBB MST Last F	NPY1R MST Last F
AR MST Last R	HBB MST Last R	NPY1R MST Last R

(Refer to materials and methods section (Table 2.1) for sequence information.)

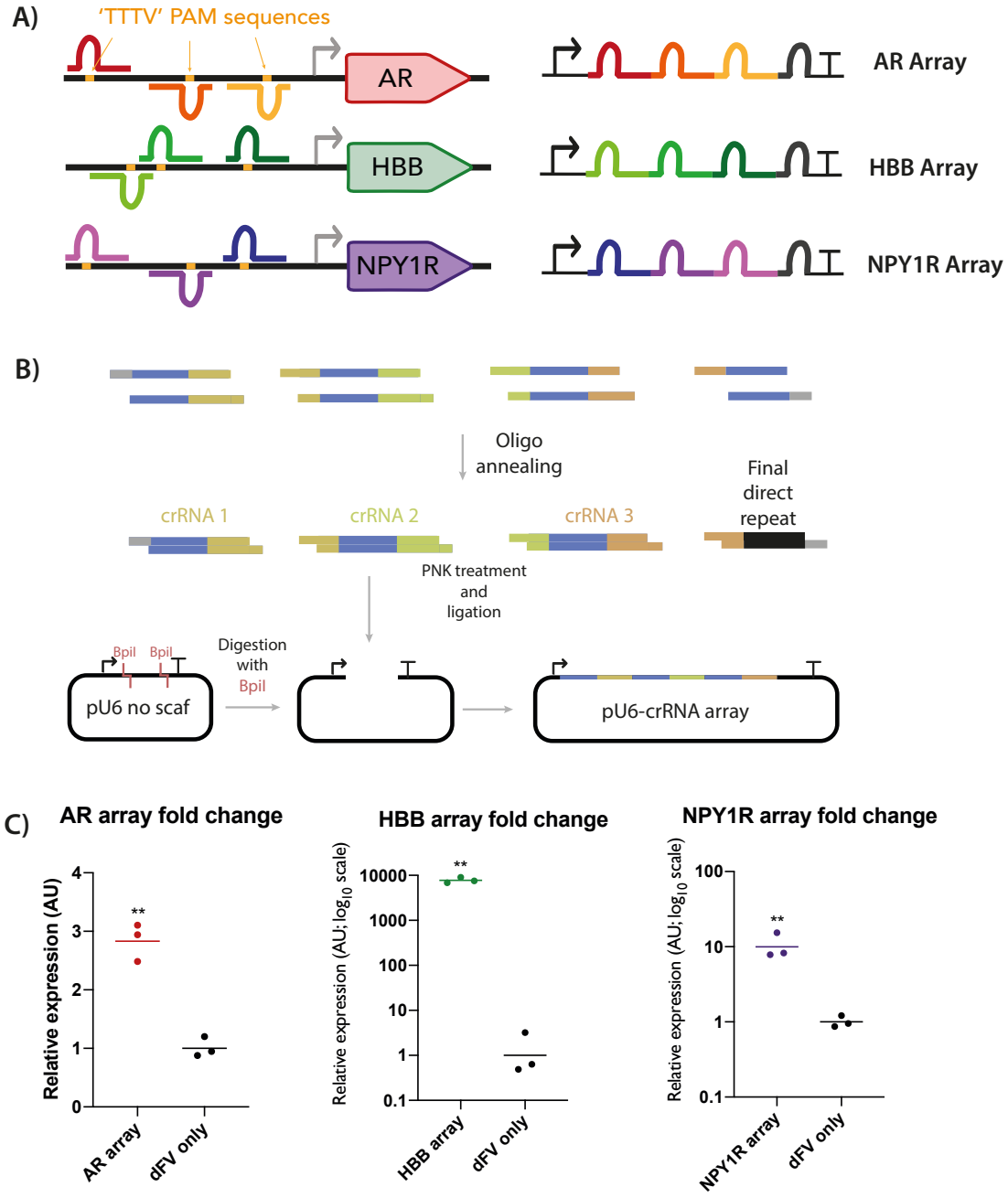

Figure 4-5 - Screening Tak et al. 3-crRNAs arrays with Fn direct repeats

A) Diagrammatic representation of the targeting of dFnCas12a-VPR to endogenous AR, HBB and NPY1R promoters in HEK293 cells, alongside the 3-crRNA arrays used for targeting of the respective promoters. B) Diagram showing the assembly strategy for construction of the crRNA arrays. C) The crRNA arrays for each gene were transfected alongside dFnCas12a-VPR into HEK293 cells. Graphs show qRT-PCR data for relative expression of the AR, HBB or NPY1R transcripts, with dFnCas12a-VPR only (dFV only) serving as a negative control, with the mean for the negative control being set to 1 and used to normalise the remaining samples (n = 3).

4.5 Screening individual crRNAs for activity against endogenous genes

Having shown that robust transactivation could be achieved for dFnCas12a-VPR when targeting endogenous genes, we next sought to address two questions. 1) Could we identify more genes that demonstrated higher fold upregulation, to improve the signal to noise ratio for subsequent experiments where reduced transactivation may be expected? 2) Could we observe significant transactivation when targeting dFnCas12a-VPR to a promoter using a single crRNA? To try and address these questions two endogenous promoters, driving expression of ASCL1 and IL1RN, were selected for targeting. These genes were chosen as they had previously shown high levels of transactivation using dCas9-VPR in the Rosser lab (Kleinjan et al., 2017). Three crRNAs were designed to target each promoter region (Figure 4.6A). The crRNAs were constrained to a window of 50 to 300 nucleotides upstream of the transcription start site (TSS) for each gene, based on work by Gilbert et al. in 2014 where they had tiled a Cas9 derived synthetic transcription factor across multiple promoter regions and observed highest activity when gRNAs targeted a window of 50 to 400 nucleotides upstream of the TSS (Gilbert et al., 2014). The TSS for each gene was identified using CAGE-Seq data curated by the FANTOM5 project (Abugessaisa et al., 2017). The crRNAs selected within this window were designed to target the characterised FnCas12a 'TTV' PAM sequence (Zetsche et al., 2015a). As subsequent work was expected to focus on dFnCas12a-VPR, a modified pU6 plasmid (U6 Fn scaf) was designed to enable cloning of only the spacer sequence into a plasmid, such that the Fn direct repeat was already present immediately downstream of the hU6 promoter (Figure 4.6B). This was achieved by digesting the U6 no scaf vector with Bpi1 and incorporating the Fn direct repeat immediately downstream of the hU6 promoter, whilst incorporating two new Bpi1 sites between the direct repeat and the terminator sequence. Two oligos (Fn DR F + Fn DR R) were annealed and ligated into the digested and gel extracted BpiI linearised U6 no scaf vector.

To assemble the three crRNAs targeting ASCL1 and IL1RN, oligos (Table 4.2) were designed to generate the spacer sequences and annealed together (Table 2.1), before ligating into the Bpi1 digested pU6 Fn direct repeat plasmid. After sequence verification, these crRNAs (all three crRNAs in a single sample or individually) were transfected alongside dFnCas12a-VPR into HEK293 cells either pooled or individually. dFnCas12a-VPR without a targeting crRNA served as negative control. Three days post transfection, total RNA was extracted from each sample and 1µg was used to synthesize cDNA using oligo dT(20) primers (further described in chapter 6). 1ul of the generated cDNA was used for qRT-PCR analysis of the target gene and normalised to reference gene GAPDH. When performing a simple ANOVA with Tukey's multiple comparisons test, we observed that a single crRNA was sufficient to enable transactivation for both endogenous genes (P = 0.02 for A2 and P = 0.006 for I2) (Figure 4.6C).

Table 4.2 - ASCL1 and IL1RN crRNA oligos

Oligo name	Oligo sequence
ASCL1 c1 F	AGATAGCTGGGTTTGTTGTTGCAG
ASCL1 c1 R	AAACCTGCAACAACAAACCCAGCT
ASCL1 c2 F	AGATCAAGGAGCGGGAGAAAGGAA
ASCL1 c2 R	AAACTTCCTTTCTCCCGCTCCTTG
ASCL1 c3 F	AGATGGGAGTGGGTGGGAGGAAGA
ASCL1 c3 R	AAACTCTTCCTCCCACCCACTCCC
IL1RN c1 F	AGATCGCAGATAAGAACCAGTTTG
IL1RN c1 R	AAACCAAACTGGTTCTTATCTGCG
IL1RN c2 F	AGATCAGGAGGGTGACTCAGGCTA
IL1RN c2 R	AAACTAGCCTGAGTCACCCTCCTG
IL1RN c3 F	AGATGCATCAAGTCAGCCATCAGC
IL1RN c3 R	AAACGCTGATGGCTGACTTGATGC

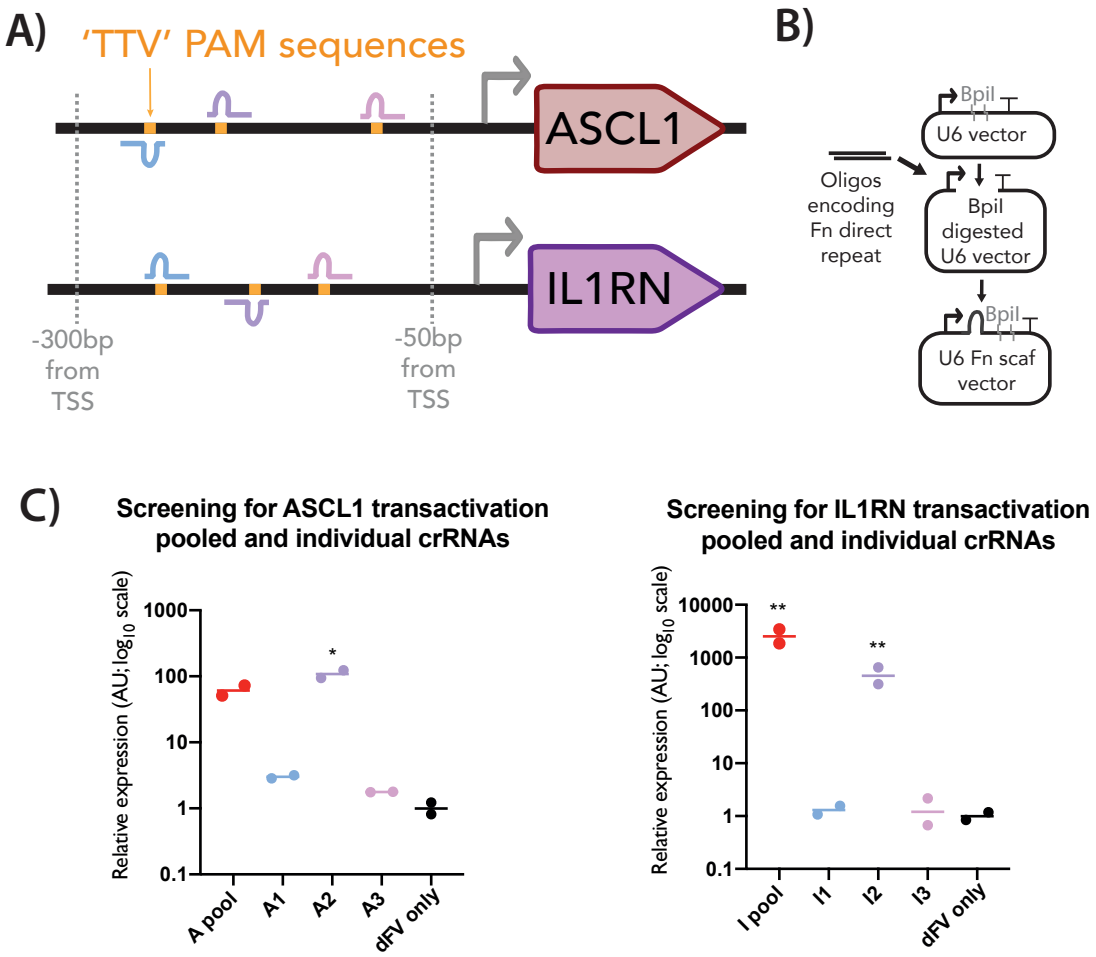

Figure 4-6 - Screening activity of individual crRNAs against endogenous genes

A) Diagrammatic representation of the targeting of dFnCas12a-VPR to endogenous promoters in HEK293 cells. Three crRNAs were designed to target a 'TTV' PAM sequence between 50 and 300bp upstream of the transcription start site for both ASCL1 and IL1RN. The transcription start site was predicted using CAGE-Seq data curated by the FANTOM5 project. B) Cloning strategy for generating U6 Fn scaf vector. The U6 vector was linearised using BpiI, before ligating with annealed oligos encoding the Fn crRNA direct repeat and new BpiI sites for subsequent cloning. C) The crRNAs for each gene were transfected individually or as a pool alongside dFnCas12a-VPR into HEK293 cells. Graphs show qRT-PCR data for relative expression of the ASCL1 or IL1RN transcripts, with dFnCas12a-VPR only (dFV only) serving as a negative control, with the mean for the negative control being set to 1 and used to normalise the remaining samples (n = 2).

4.6 Identifying multiple active crRNAs targeting 3 genes

Having demonstrated robust transactivation could be achieved with a single crRNA, we sought to identify two or more active crRNAs across three different genes. This would enable a number of subsequent experiments and analysis to be performed, allowing the exploration of the hypothesis that additive or synergistic up-regulation can be achieved when targeting multiple active crRNAs to a single promoter. We chose HBB as a third gene to target, as we had previously observed robust transactivation could be achieved when targeting its promoter with dFnCas12a-VPR (Figure 4.5B). As previously we had been able to identify an active crRNA when screening three unique crRNAs, for targeting both the ASCL1 and IL1RN promoter regions, we screened 6 crRNAs for each of the three promoters (Figure 4.7A), expecting at least 2 active crRNAs to be identified for each promoter. As three crRNAs were already available for targeting ASCL1 and IL1RN, a further three crRNAs were designed using the previously described design constraints. Six unique crRNAs were also designed for the targeting of the HBB promoter, employing the same design constraints (Figure 4.7A). The crRNAs were generated (Table 4.3) as described, by annealing oligos and ligating into BpiI digested pU6 Fn direct repeat plasmid (Materials and methods 2.1.1 and 2.1.2). After sequence verification, each of these crRNAs were individually transfected into HEK293 cells along with dFnCas12a-VPR. Three days after transfection, RNA was extracted for cDNA synthesis and qRT-PCR was performed.

Table 4.3 - Oligos for generating ASCL1, HBB and IL1RN crRNAs

Oligo name	Oligo sequence
Fn HBB g1 F 2	CACCGTAATTTCTACTGTTGTAGATTACTGATGGTATGGGGCCAA
Fn HBB g1 R 2	AAACTTGGCCCCATACCATCAGTAATCTACAACAGTAGAAATTAC
Fn HBB g2 F 2	CACCGTAATTTCTACTGTTGTAGATAAGTCCAACTCCTAAGCCAG
Fn HBB g2 R 2	AAACCTGGCTTAGGAGTTGGACTTATCTACAACAGTAGAAATTAC
Fn HBB g3 F 2	CACCGTAATTTCTACTGTTGTAGATCAAGTGTATTTACGTAATAT
Fn HBB g3 R 2	AAACATATTACGTAAATACACTTGATCTACAACAGTAGAAATTAC
ASCL1 crRNA 4 F	AGATTTGTTGCAGTGCGTGCGCCT
ASCL1 crRNA 4 R	AAACAGGCGCACGCACTGCAACAA
ASCL1 crRNA 5 F	AGATtcccgCTCCTTGCAAACTCT
ASCL1 crRNA 5 R	AAACAGAGTTTGCAAGGAGcggga
ASCL1 crRNA 6 F	AGATctttctcccgCTCCTTGCAA
ASCL1 crRNA 6 R	AAACTTGCAAGGAGcgggagaaag
HBB crRNA 4 F	AGATGTAGCAATTTGTACTGATGG
HBB crRNA 4 R	AAACCCATCAGTACAAATTGCTAC
HBB crRNA 5 F	AGATGAGGGAGGGCTGAGGGTTTG
HBB crRNA 5 R	AAACCAAACCCTCAGCCCTCCCTC
IL1RN crRNA 4 F	AGATTCTGCATGTGACCTCCCATC
IL1RN crRNA 4 R	AAACGATGGGAGGTCACATGCAGA
IL1RN crRNA 5 F	AGATGTTTCTGCTAGCCTGAGTCA
IL1RN crRNA 5 R	AAACTGACTCAGGCTAGCAGAAAC
IL1RN crRNA 6 F	AGATGCCAGCATGAGGAGATGGGC
IL1RN crRNA 6 R	AAACGCCCATCTCCTCATGCTGGC

When performing a simple ANOVA with Dunnett's multiple comparisons test, identified 5 crRNAs across the three different genes that displayed significant transactivation (Figure 4.7B): A2 (P = 0.0003) for ASCL1; H1 (P = 0.003), H2 (P = 0.001), H4 (P = 0.03) for HBB and I2 (P = 0.0002) for IL1RN. Two crRNAs A4 targeting ASCL1 and I5 targeting IL1RN may also possess some activity, however their activity was non-significant (P = 0.18 and P = 0.09 respectively). Therefore, to determine combined activity of multiple crRNAs, the A2 + A4, I2

+ I5 and H1 + H2 pairs were chosen. We also screened the next most active crRNAs where the data suggested the potential for activity, including H4 + H5 and I4 + I6 into subsequent screening to determine whether crRNAs with low or non-significant activity would show enhanced or even synergistic activity upon co-transfection.

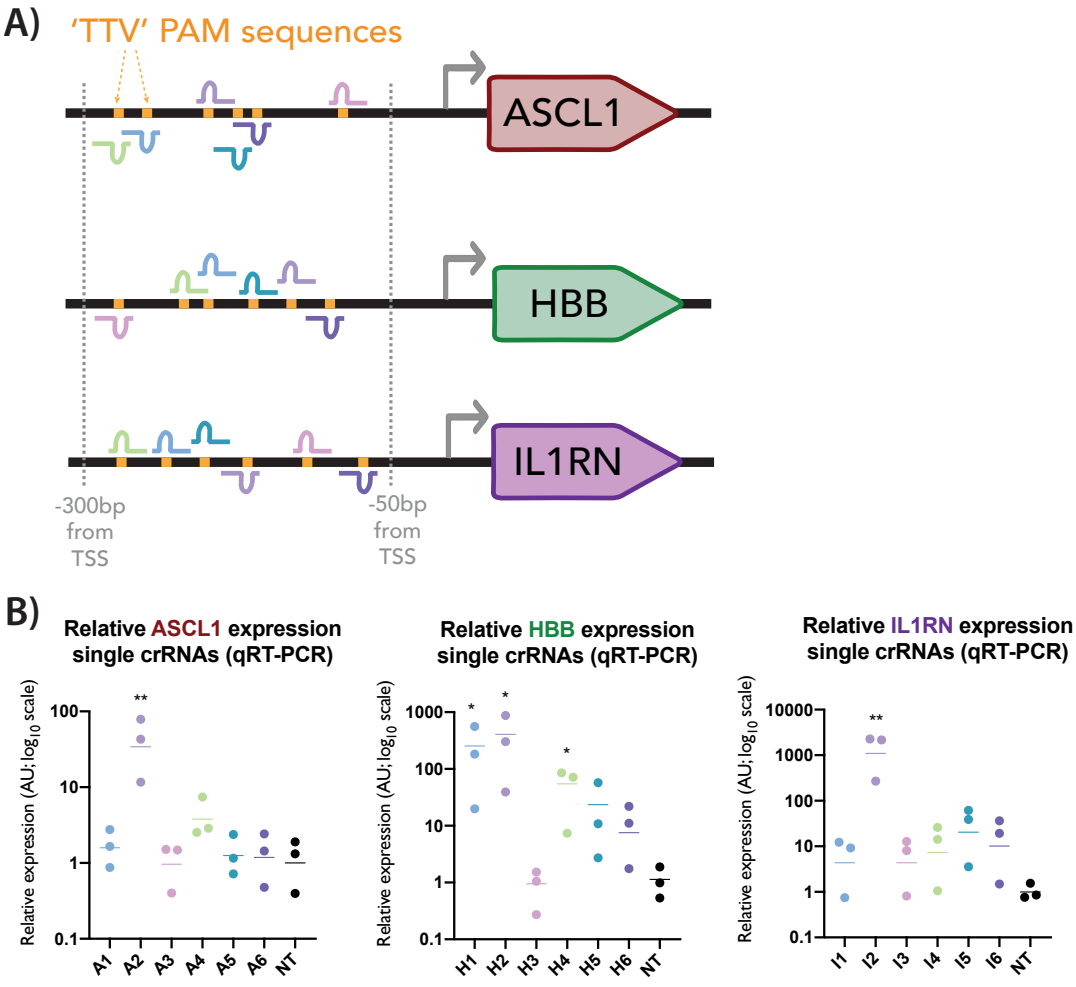

Figure 4-7 - Identifying active crRNAs across multiple genes

A) Diagrammatic representation of the targeting dFnCas12a-VPR to three endogenous promoters in HEK293 cells. Alongside the 3 previously screened crRNAs targeting ASCL1 and IL1RN, 3 further crRNAs targeting the 'TTV' PAM sequence were designed and generated for each gene. 6 crRNAs were designed and generated for targeting the HBB promoter, following the same design constraints described for ASCL1 and IL1RN. B) The crRNAs for each gene were transfected individually with dFnCas12a-VPR into HEK293 cells. qRT-PCR analysis was performed on cells collected after 3 days incubation. Graphs show qRT-PCR data for relative expression of the ASCL1, HBB or IL1RN transcripts, with dFnCas12a-VPR delivered with a non-targeting crRNA serving as negative control (n = 3). The mean for each negative control was set to 1 and used to normalise the remaining samples.

4.7 Delivering crRNA pairs improves transactivation

We used our crRNA pairs to screen for enhanced transactivation in HEK293 cells, compared to the individual crRNAs (Figure 4.8A). A non-targeting crRNA (the 'alt PAM Fn Luc' crRNA sequence used for targeting the luciferase reporter vector in Figure 4.2) served as negative control. All crRNAs were co-transfected with dFnCas12a-VPR. Three days post transfection RNA was extracted, cDNA generated and qRT-PCR performed as previously described. When a Welch's t test was performed to compare the activity of the crRNA pairs against the most active individual crRNA, for all three genes at least one of the pairs showed significantly higher activity than the most active single crRNA (Figure 4.8B). A2 + A4 co-transfected showed significantly higher transactivation than the most active individual crRNA - A2 ($P = 0.001$), H1 + H2 co-transfected showed significantly higher transactivation than the most active single crRNA - H2 ($P = 0.02$), and I4 + I6 showed significantly higher transactivation than the most active single crRNA - I6 ($P = 0.02$). An interesting anomaly was a non-significant ($P = 0.12$) reduction in activity observed when I2 + I5 were co-delivered compared to the activity of I2 delivered alone. This result can be explained by the fact that the two crRNAs target opposite strands of the genome, with 12 nucleotides of overlap for the respective spacer sequences. This presents two models, the first being direct hybridisation between the crRNAs minimising the pool of active crRNAs available for targeting. The second model is based on competitive inhibition, as only one of two crRNA/dCas12a-VPR complexes can stably bind the DNA at any time, leading to the relative expression being lower than that observed for the most active single crRNA (I2). Contrary to our previous results, we also were able to see significant transactivation for I4 ($P = 0.003$) and I6 ($P = 0.002$), which may in part be explained by the fact that previously one of the three repeats showed significantly lower transactivation across all samples (Figure 4.7C) and when removed, significant transactivation was observed for both I4 ($P = 0.02$) and I6 ($P = 0.02$).

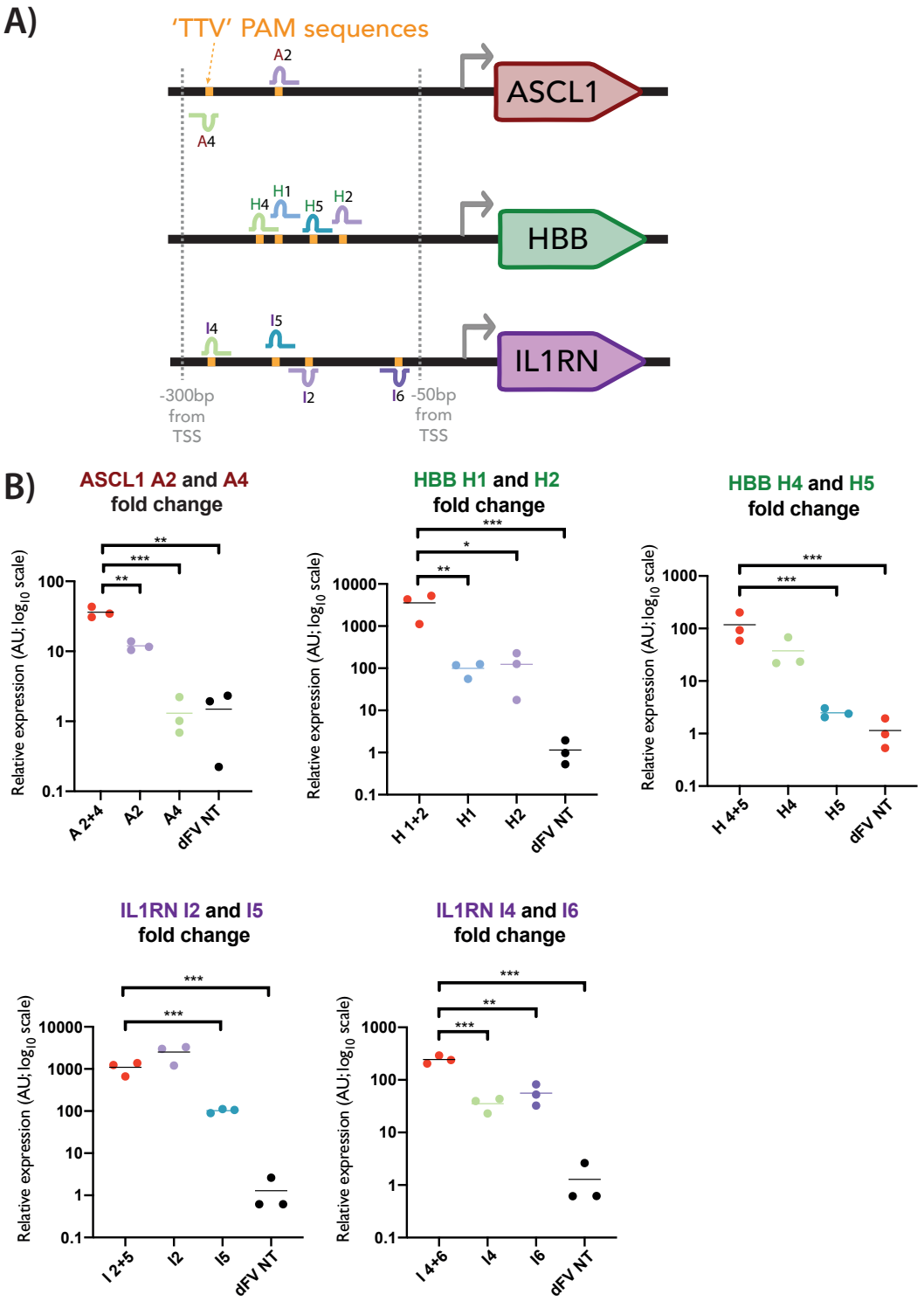

Figure 4-8 - Screening activity of pooled vs. individual crRNAs

A) Diagrammatic representation of the crRNAs used for targeting dFnCas12a-VPR to the ASCL1, HBB and IL1RN promoters. B) Graphs show qRT-PCR data for the relative expression of the ASCL1, HBB or IL1RN transcripts (n = 3). Two individual or pooled crRNAs were delivered to HEK293 cells alongside a non-targeting crRNA serving as negative control. The mean for each negative control was set to 1 and used to normalise the remaining samples.

4.8 Screening pooled crRNA for synergy

An important question when considering the delivery of multiple crRNAs to target multiple loci is whether a focus should be given to utilising multiple crRNAs to each promoter or to focus on single crRNAs for a wider range of target promoters. An important consideration is the level of impact observed when multiple active crRNAs are targeted to the same promoter, in particular whether we can observe a synergistic (greater than additive) increase to transactivation when more than one crRNA are delivered.

To assess for synergy we can compare the mean and standard deviation of the experimental result when both crRNAs were delivered to that of a hypothetical additive distribution. The hypothetical distribution can be inferred from the means and standard deviations of the experimental results when each crRNA is delivered separately. This would allow us to draw an accurate assessment of whether the experimental distribution we were sampling from was distinct from the hypothetical distribution and therefore whether synergy was observed (Figure 5.1A).

All statistics were performed using $\Delta\Delta$Ct values, which are assumed to be normally distributed (as opposed to the actual fold change values, which are log-normally distributed due to the $2^{-(\Delta\Delta Ct)}$ transformation). The additive condition is defined as the sum of the log fold changes (or $\Delta\Delta$Ct values) when the given crRNAs are introduced independently, compared to a non-targeting crRNA. The mean of the hypothetical additive distribution can be calculated by adding together the fold

changes observed for the each individual crRNA and taking log base 2 of this result. The variance could be calculated using the formula described in Materials and Methods 2.3.4, using the means and variances from the individual crRNA results. These calculated means and standard deviations for the hypothetical additive values for all 5 crRNA pairs were then used to test for significant separation compared to the results obtained experimentally when the crRNA pairs were co-delivered. The results are also visually represented in Figure 5.1B.

After calculating the means for the predicted additive distributions, we were able to see that the mean of the experimentally observed pooled conditions were higher than the mean of the hypothetical additive for 4 of the 5 crRNA pairs, with the only exception being the I2+I5 condition.

To assess the significance of separation between the hypothetical additive distributions and experimentally observed pooled distributions, we used a Welch's t test. For all 5 pairs tested, we observed no significant synergy; A 2+4 ($P = 0.13$) H 1+2 ($P = 0.1$), H 4+5 ($P = 0.17$), I 2+5 ($P = 0.09$) and I 4+6 ($P = 0.44$). Graphs showing the $\Delta\Delta$Ct values for the crRNA pairs are shown in (Figure 4.9B). The mean and standard deviation for the hypothetical distribution is plotted alongside the experimental data for the pooled, individual and negative control conditions.

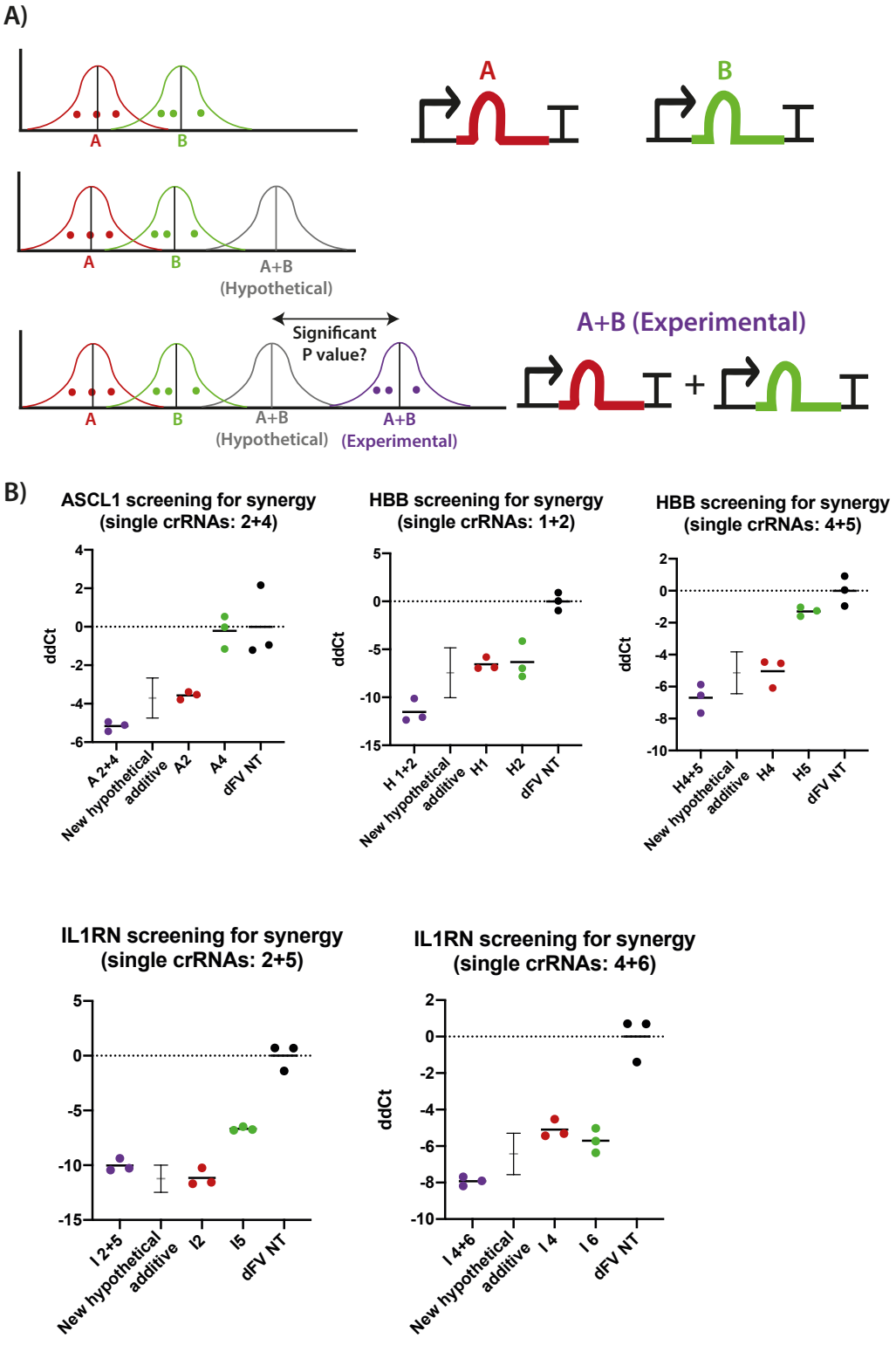

Figure 4-9 - Screening for synergistic activity of pooled crRNAs

A) Diagrammatic representation of the generation of a hypothetical additive distribution from two individual experimentally derived distributions, alongside the subsequent comparison between the hypothetical additive condition and the experimental combined condition. B) Graphs showing the ΔΔCt for the experimental individual and pooled crRNA conditions alongside the hypothetical additive distribution with error bars showing standard deviation.

4.9 Conclusion and Discussion

So far, we have demonstrated that synthetic transcription factors can be generated from Cas12a variants derived from three different species, confirming observations seen for dAsCas12a-VPR and dLbCas12a-VPR (Tak et al., 2017). We have presented a novel finding, demonstrate robust transactivation can be achieved by the dFnCas12a-VPR variant, which possesses the shortest described PAM sequence for the Cas12a family 'TTV' (Leenay et al., 2016). We have explored orthogonality between the three different dCas12a-VPR variants, observing evidence of orthogonality between dFnCas12a-VPR and dLbCas12a-VPR. Further characterisation of dFnCas12a-VPR for three different genes, showed that a single crRNA is sufficient to enable transactivation.

Finally, we demonstrate that similar to results observed for Cas9-derived synthetic transcription factors (Maeder et al., 2013a), it is possible to observe greater transactivation of a target gene when multiple active crRNAs are delivered to target the same promoter. However, when testing for synergy, we do not observe significantly greater than additive transactivation.

The desire for more targeting crRNAs to enable improved transactivation of individual genes, is naturally supported by the Cas12a system, as the Cas12a protein has been shown to process its own crRNA array (Fonfara et al., 2016). This means that longer crRNA arrays expressed from on a single transcript should in principle enable strong transactivation to be achieved for multiple genes, something we explore further in the subsequent chapter.

Chapter 5 Synergistic activation and crRNA arrays

5.1 Introduction

Work with Cas9 derived synthetic transcription factors (Maeder et al., 2013a), alongside more recent work with dLbCas12a-VPR (Tak et al., 2017) has shown that increased transactivation of targeted loci can be achieved when multiple synthetic transcription factors are targeted to the same loci. Work with dLbCas12a-VPR has also highlighted the capability of crRNA arrays to enable targeting of multiple crRNA to different loci from the same transcript. We wanted to test whether these observations were also seen for dFnCas12a-VPR. Furthermore, we wished to test the limitations of crRNA arrays, exploring the capacities of longer arrays. Finally, we sought to pursue novel applications of dFnCas12a-VPR, focussing on maximising the strengths of the highly dense targeting capabilities. To achieve this we built on previous work with Cas9, where variants of the wild-type Cas9 (Wright et al., 2015) and even a Cas9 based synthetic transcription factor (Zetsche et al., 2015c) were split into 2 fragments and shown to preserve activity when co-expressed. As such we sought to explore the generation of split dFnCas12a-VPR, to in particular enable AAV viral delivery, for future applications including gene therapy.

5.2 Screening crRNA arrays for activity

Having screened for activity of pooled crRNAs, we next explored the capacity of crRNA arrays to enable transactivation of target genes. We proceeded to design

and generate crRNA arrays for 4 of the 5 crRNA pairs previously screened (A 2+4, H 1+2, H 4+5 and I 4+6). The crRNA arrays were designed to incorporate an extra direct repeat at the 3' of the second crRNA to recapitulate the designs previously utilised by Tak et al (Figure 5.1A) (Tak et al., 2017). The crRNA arrays were generated by annealing and ligating three pairs of oligos into the digested pU6 no scaffold vector (Table 2.1) (Materials and Methods 2.1.1 and 2.1.2). After sequence verification, the constructed arrays were transfected into HEK293 cells as previously described, alongside the crRNAs pairs delivered together or individually. dFnCas12a-VPR delivered with a non-targeting crRNA was used as a negative control. 3 days post transfection, RNA was extracted, cDNA generated and qRT-PCR performed as previously described. The first key observation, was that all 4 crRNA arrays tested, across all three genes, showed highly significant transactivation compared to the negative control when a Tukey's multiple comparisons test was performed; A array (2+4) ($P < 0.0001$), H array (1+2) ($P < 0.0001$), H array (4+5) ($P < 0.0001$) and I array (4+6) ($P = 0.005$) (Figure 5.1B). Of interest, we consistently saw across all arrays tested, that the mean for each array was higher than the mean for the pooled condition. However, when employing Tukey's multiple comparisons test to compare the activity of the arrays to the crRNA pairs, only the HBB 4+5 array showed significantly higher activity than the corresponding crRNA pair ($P = 0.02$). Finally, we observed significantly higher transactivation for each array condition compared to the highest expressing individual crRNA for 2 of the 4 arrays tested; HBB 12 array (H1+H2) ($P < 0.0001$) and HBB 45 array (H4+H5) ($P = 0.0002$).

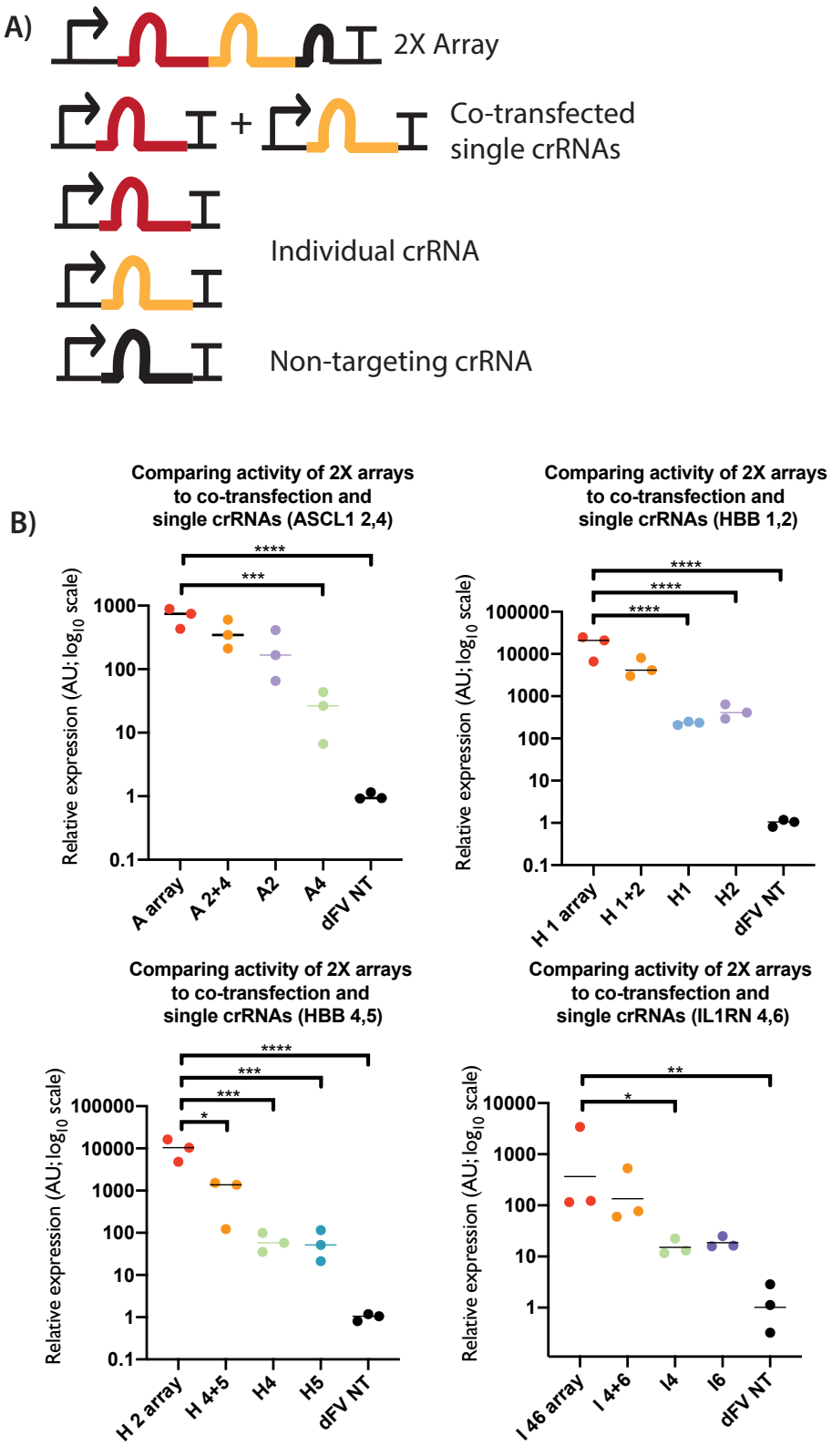

Figure 5-1 - Screening crRNA arrays for activity

A) Diagrammatic representation of the different constructs tested side by side in the subsequent qRT-PCR analysis. B) Graphs showing the fold change in expression of the three genes (ASCL1, HBB and IL1RN), comparing the array, combined crRNAs and individual crRNAs to the negative control (dFnCas12a-VPR delivered with a non-targeting crRNA) normalised to 1 (n = 3).

5.3 Screening crRNA arrays for synergy

Having observed robust transactivation (Figure 5.2A) could be achieved using these short crRNA arrays, we next wished to test whether the arrays demonstrated synergistic transactivation. As described for testing synergy with pooled crRNAs, we calculated a 'hypothetical additive distribution' by combining means and variances of the two individual crRNA distributions as previously described (FIGURE 5.2A). To assess the significance of separation between the hypothetical additive distributions and experimentally observed array distributions, we used a Welch's t test. Two of the arrays tested demonstrated synergistic transactivation; H12 array (H1 + H2) (P = 0.005) and H45 array (P = 0.005). Graphs showing the $\Delta\Delta$Ct values for the crRNA pairs are shown in (Figure 5.2B). The mean and standard deviation for the hypothetical distribution is plotted alongside the experimental data for the pooled, individual and negative control conditions.

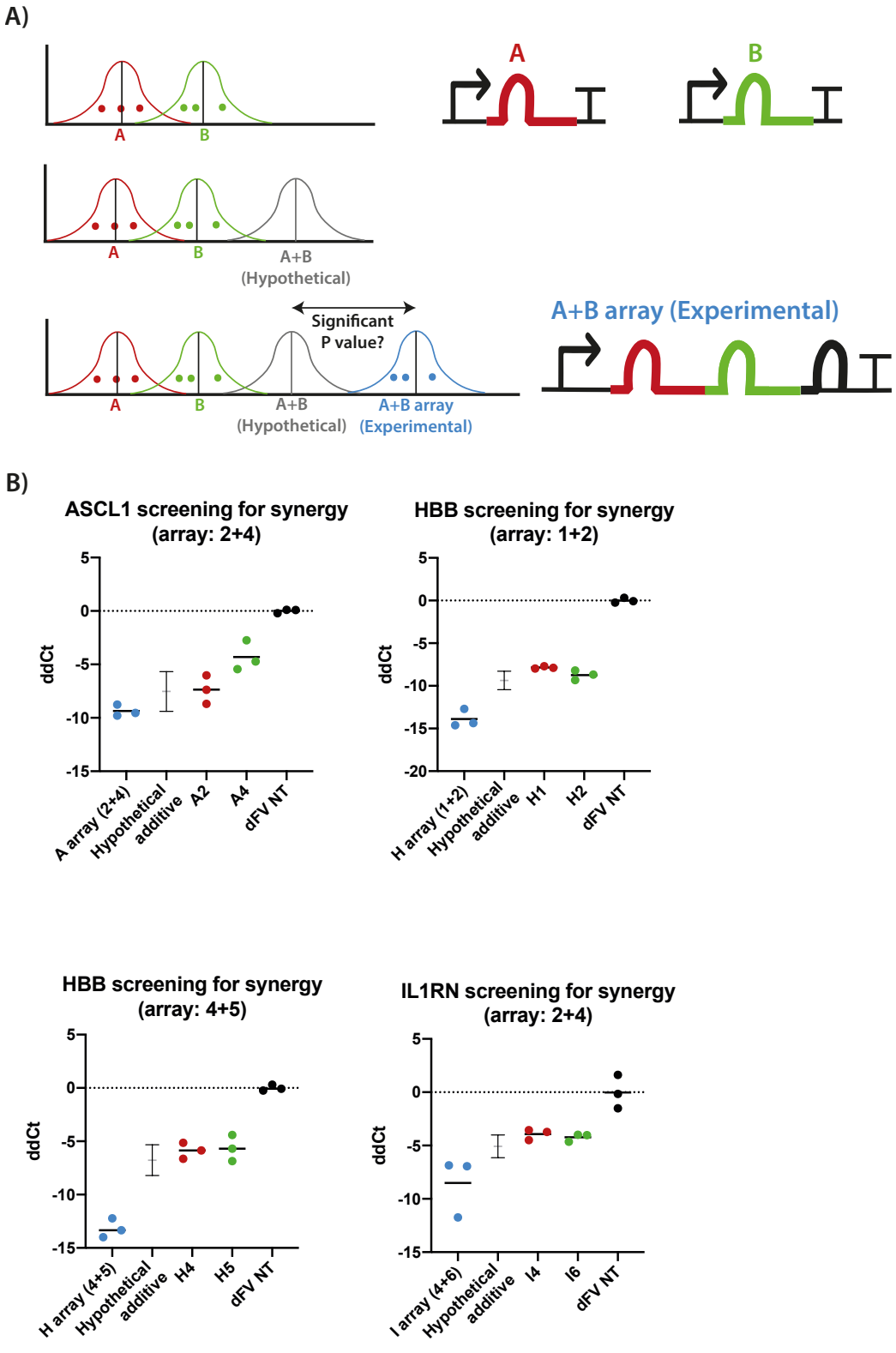

Figure 5-2 - Screening for synergistic activity of crRNA arrays

A) Diagrammatic representation of the generation of a hypothetical additive distribution from two individual experimentally derived distributions, alongside the subsequent comparison between the hypothetical additive condition and the experimental array condition. B) Graphs showing the ΔΔCt for the experimental individual crRNA and array conditions alongside the hypothetical additive distribution with error bars showing standard deviation.

5.4 Screening 6-crRNA arrays for multiplexing

Having observed that 2-crRNA arrays could enable robust transactivation of target genes, we next sought to test whether they could enable multiplexed transactivation within 6-crRNA arrays, with all three pairs being expressed from a 6-crRNA array. We also sought to test whether there was a position dependent activity of the crRNA pairs depending on their position within the crRNA array.

Six different 6-crRNA arrays were designed, such that the crRNAs in the first two positions would target one gene, the crRNAs in the next two positions would target a second gene and the crRNAs in the final two positions would target a third gene. The most active 2-crRNA arrays for ASCL1, HBB and IL1RN were utilised (A array (A2+A4), H12 array (H1+H2) and I46 array (I4 + I6)) (Figure 5.3A). All possible combinations were designed (Figure 5.3B). The 6-crRNA arrays were generated by annealing oligo pairs (Table 5.1) and ligating into a BpiI digested U6 vector (Materials and Methods 2.1.1 and 2.1.2). After sequence verification of the constructs, they were transfected alongside the following positive control 2-crRNA arrays; A array (A2+4), H12 array (H1+H2) and I46 array (I4+I6). A non-targeting crRNA served as a negative control. All the preceding constructs were delivered alongside dFnCas12a-VPR, with dFnCas12a-VPR delivered with a non-targeting crRNA serving as a negative control. Three days post transfection, RNA was extracted, cDNA generated and qRT-PCR performed as previously described (Figure 5.3C).

Table 5.1 - Oligo pairs used for constructing 6x crRNA for multiplexing

AHI array	AIH array	HAI array	HIA array	IAH array	IHA array
A2 first F	A2 first F	H1 first F	H1 first F	I4 first F	I4 first F
A2 first R	A2 first R	H1 first R	H1 first R	I4 first R	I4 first R
A24 F	A24 F	H1H2 F	H1H2 F	I4I6 F	I4I6 F
A24 R	A24 R	H1H2 R	H1H2 R	I4I6 R	I4I6 R
A4H1 F	A4I4 F	H2A2 F	H2I4 F	I6A2 F	I6H1 F
A4H1 R	A4I4 R	H2A2 R	H2I4 R	I6A2 R	I6H1 R
H1H2 F	I4I6 F	A24 F	I4I6 F	A24 F	H1H2 F
H1H2 R	I4I6 R	A24 R	I4I6 R	A24 R	H1H2 R
H2I4 F	I6H1 F	A4I4 F	I6A2 F	A4H1 F	H2A2 F
H2I4 R	I6H1 R	A4I4 R	I6A2 R	A4H1 R	H2A2 R
I4I6 F	H1H2 F	I4I6 F	A24 F	H1H2 F	A24 F
I4I6 R	H1H2 R	I4I6 R	A24 R	H1H2 R	A24 R
I6 last F	H2 last F	I6 last F	A4 Last F	H2 last F	A4 Last F
I6 last R	H2 last R	I6 last R	A4 Last R	H2 last R	A4 Last R

(Refer to materials and methods section (Table 2.1) for sequence information.)

The first key observation was that for all 6-crRNA arrays screened we observed highly significant transactivation above the negative control when Tukey's multiple comparison test was performed for all three genes ($P < 0.0001$ for all arrays for all three genes) (Figure 5.3C). A second interesting observation was seen when comparing the positive controls to the arrays where their respective targeting crRNAs were in position 5 and 6. For all three genes, we saw that the positive control showed significantly higher activation than at least one of these two arrays. For ASCL1, when comparing the A array to IHA we observed a significantly lower activity for IHA ($P = 0.02$), for HBB when comparing the H12 array to AIH and IAH we also observed significantly lower activity for the 6-crRNA arrays ($P = 0.02$ and $P = 0.01$ respectively). Finally, for IL1RN when comparing the I46 array to AHI and HAI

we also observed significantly lower activity for the 6-crRNA arrays (P = 0.0001 and P = 0.0002 respectively). These findings alongside the consistent visual downward trend observed when the fold activation of arrays were ordered by position of the active crRNAs within the 6-crRNA arrays (Figure 5.3B) led to the hypothesis that there may be a decrease in activity when crRNAs are in 3' positions within longer arrays.

To test this hypothesis, after ordering the cycle threshold results based upon the position of the targeting crRNAs within each array (e.g. AHI, AIH, HAI, IAH, HIA, IHA for ASCL1) a statistical test for linear trend was performed, with a significant downwards trend observed for ASCL1 (P = 0.0001), HBB (P = 0.001) and IL1RN (P < 0.0001). When analysed together, the results observed for all three genes suggested that whilst active throughout the 6 crRNA array, crRNAs appeared to show reduced activities when closer to the 3' of the crRNA array.

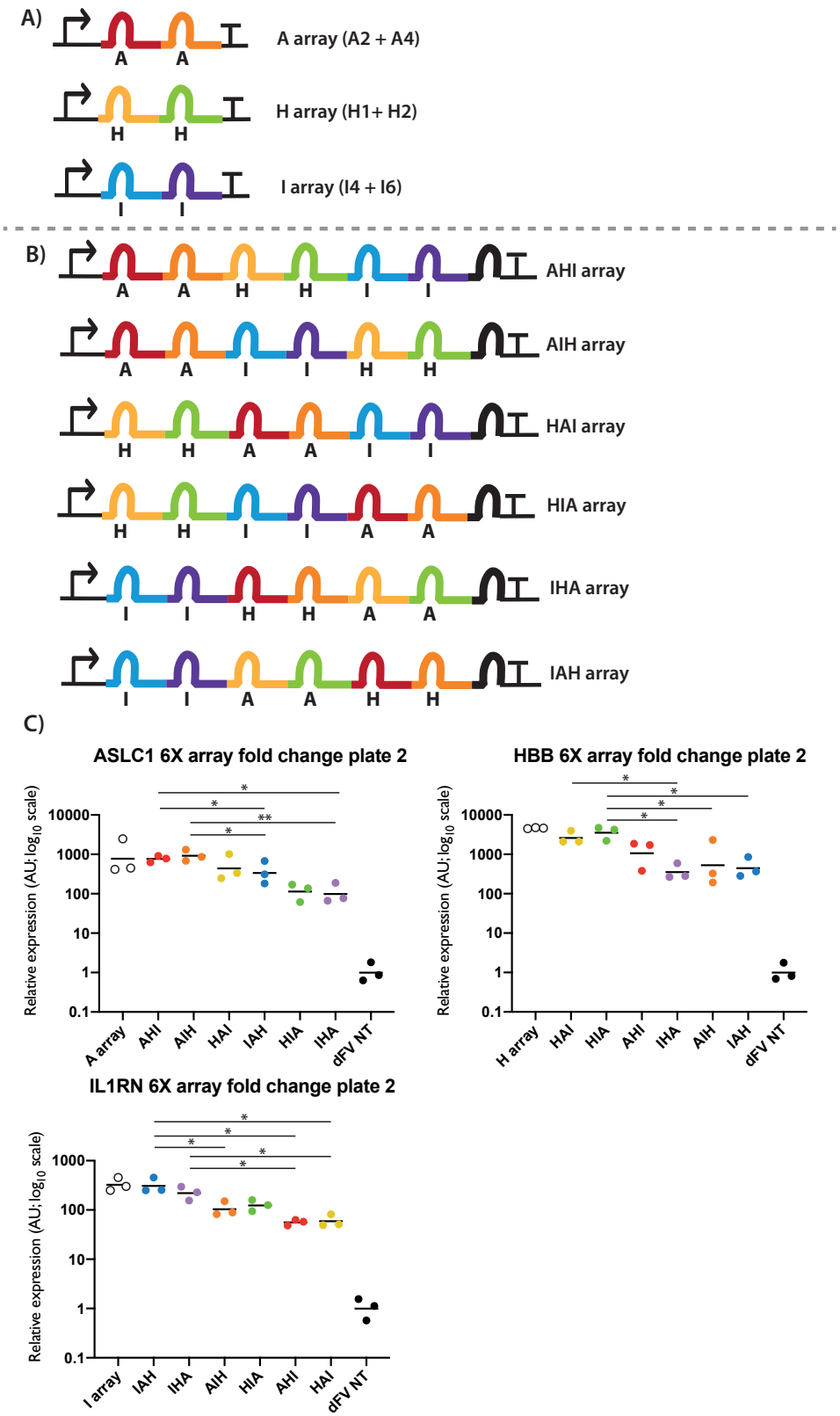

Figure 5-3 - Testing for multiplexing using 6x crRNA arrays

A) Diagrammatic representation of the different positive control arrays. B) Diagrammatic representation of 6X crRNA arrays screened by qRT-PCR C) Graphs showing qRT-PCR data for the relative expression of ASCL1, HBB or IL1RN transcripts. The mean for the respective negative controls (dFV NT) was set to 1 and used to normalise the remaining samples (n = 3).

5.5 Screening 6-crRNA array for position dependent activity of individual crRNA

Having observed position dependent activity for crRNA pairs within a 6-crRNA array we next wanted to further dissect the role of position on crRNA activity. In particular we sought to test if this trend was also observed for individual crRNA targeting, by screening all possible positions of a single targeting crRNA within 6-crRNA arrays. To test we generated a new set of 6-crRNA arrays, however this time only a single active crRNA 'A2', was present within the 6-crRNA arrays, with each of 6 arrays to be screened having A2 in a different position (1 through 6). For the remaining 5 positions, non-targeting crRNA were incorporated (Figure 5.4A). The spacer sequences for these non-targeting crRNAs were designed by first generating 20bp random DNA sequences using an online tool and subsequently performing nucleotide blast alignments against the human genome, selecting sequences where no perfect matches were observed. The 6-crRNA arrays were generated as previously described, annealing oligo pairs (Table 5.2) and ligating into a BpiI digested U6 vector (Materials and Methods 2.1.1 and 2.1.2). After sequence verification of the constructs, they were transfected alongside A array (A2 + 4) serving as a positive control. A non-targeting crRNA served as a negative control. All the preceding constructs were delivered alongside dFnCas12a-VPR. 3 days post transfection, RNA was extracted, cDNA generated and qRT-PCR performed as previously described.

The first key observation was that, as before each of the 6 crRNAs were able to show highly significant transactivation of ASCL1 when a Tukey's multiple comparisons test was performed ($P < 0.0001$ for all 6-crRNA arrays) (Figure 5.4B). Furthermore we were also able to observe that when comparing the distributions for all 6 6-crRNA arrays, only Pos1 (active A2 crRNA in the most 5' position) and Pos6 (active A2 crRNA in the most 3' position) showed significant differences between one another, with the transactivation for Pos6 being significantly lower than Pos1 ($P = 0.028$). Finally, when performing a statistical test for linear trend, we were able to observe a significant negative trend as the active crRNA moved from the most 5' position towards the most 3' position of the array ($P = 0.007$). Taken together these results provide strong evidence for position dependent activity for 6-crRNA arrays.

Table 5.2 - Oligos used for constructing position dependent ASCL1 6x array

Pos1 array	**Pos2 array**	**Pos3 array**	**Pos4 array**	**Pos5 array**	**Pos6 array**
A 1st 1 F	NT 1st F	NT 1st F	NT 1st F	NT 1st F	NT 1st F
A 1st 1 R	NT 1st R	NT 1st R	NT 1st R	NT 1st R	NT 1st R
A 1st 2 F	A 2nd 1 F	NT 12 F	NT 12 F	NT 12 F	NT 12 F
A 1st 2 R	A 2nd 1 R	NT 12 R	NT 12 R	NT 12 R	NT 12 R
NT 23 F	A 2nd 2 F	A 3rd 1 F	NT 23 F	NT 23 F	NT 23 F
NT 23 R	A 2nd 2 R	A 3rd 1 R	NT 23 R	NT 23 R	NT 23 R
NT 34 F	NT 34 F	A 3rd 2 F	A 4th 1 F	NT 34 F	NT 34 F
NT 34 R	NT 34 R	A 3rd 2 R	A 4th 1 R	NT 34 R	NT 34 R
NT 45 F	NT 45 F	NT 45 F	A 4th 2 F	A 5th 1 F	NT 45 F
NT 45 R	NT 45 R	NT 45 R	A 4th 2 R	A 5th 1 R	NT 45 R
NT 56 F	NT 56 F	NT 56 F	NT 56 F	A 5th 2 F	A 6th 1 F
NT 56 R	NT 56 R	NT 56 R	NT 56 R	A 5th 2 R	A 6th 1 R
NT last F	NT last F	NT last F	NT last F	NT last F	A 6th 2 F
NT last R	NT last R	NT last R	NT last R	NT last R	A 6th 2 R

(Refer to materials and methods section (Table 2.1) for sequence information.)

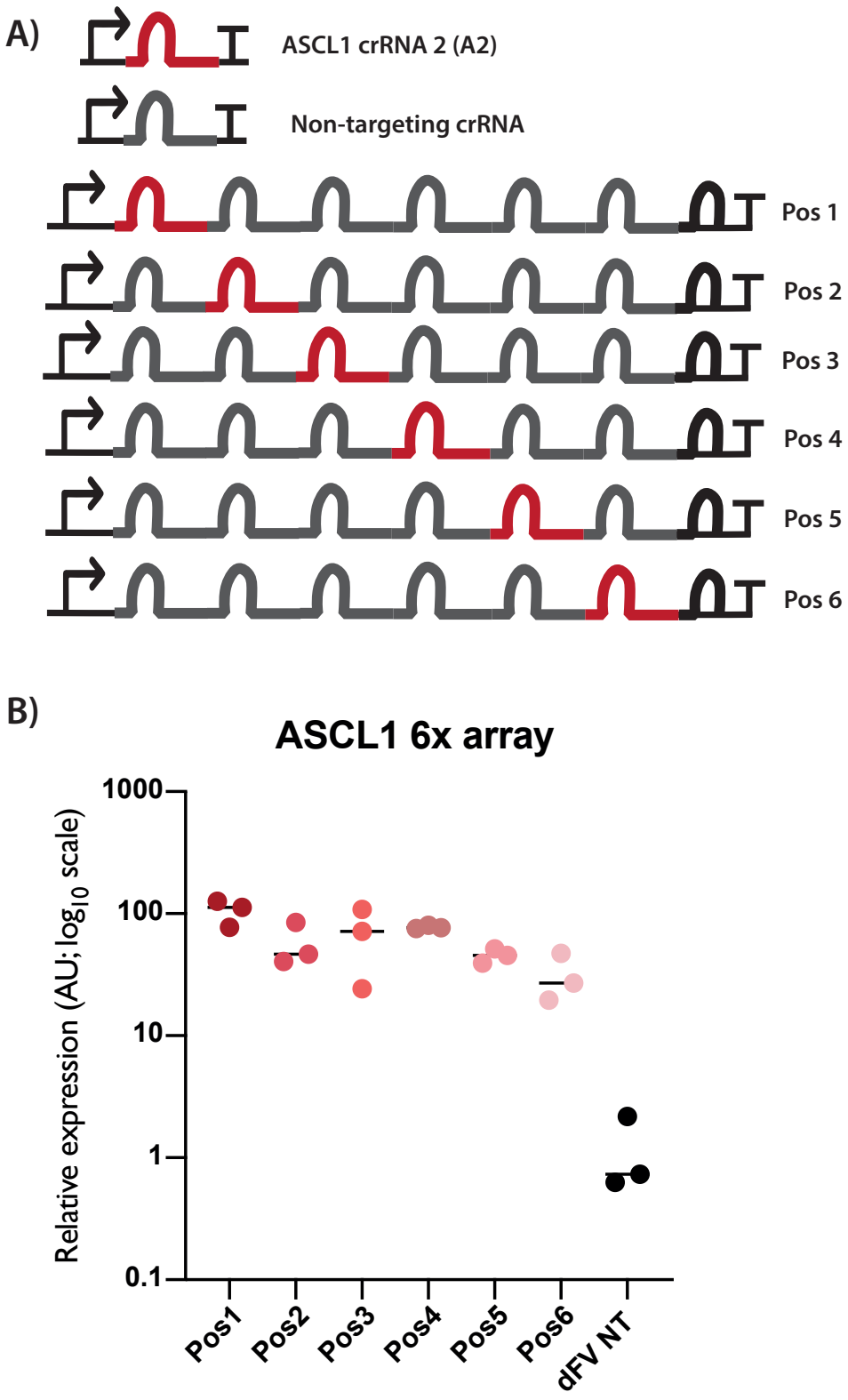

Figure 5-4 - Testing for position dependent activity within 6x crRNA array

A) Diagrammatic representation of the 6X crRNA arrays for screening for the impact of position on activity using qRT-PCR B) Graph showing qRT-PCR data for the relative expression of ASCL1 transcripts. The mean for the negative control (dFV NT) was set to 1 and used to normalise the remaining samples (n = 3).

5.6 Testing mutated direct repeat sequence for ameliorating position dependent activity

As we had identified a position dependent effect for crRNA activity within crRNA arrays, we next wanted to ascertain the cause and look for strategies to prevent reduction in activity at the 3' positions. One likely hypothesis included poor processivity of transcription, leading to premature termination of transcripts, which could potentially be explained by the presence of a premature transcriptional termination sequence within the crRNA array. We looked at the most relevant sequence that was consistently found in all arrays tested, the direct repeat, and observed that a non-canonical pol3 terminator sequence could be found (Orioli et al., 2011). When performing transcription analysis of non-canonical terminators, Orioli et al. screened the capacity of a number of putative terminators to induce premature termination of a transcript before a downstream canonical 'TTTTT' terminator sequence. Of particular interest they observed a ~5% reduction in read through of transcription for the sequence TTTCT compared to their negative control sequence 'TTT'. As each of the 7 direct repeats within the 6-crRNA array possesses the 'TTTCT' sequence, the drop-off in activity may be due to the presence of this sequence, as studies have shown that activity of wild-type Cas12a has been shown to decrease as the concentration of crRNA decreases (Kallimasioti-Pazi et al., 2018).

We sought to explore whether it would be possible to screen crRNA arrays that did not possess the non-canonical termination sequence TTTCT identified by Orioli et al. In their original work characterising Cas12a, Zetsche et al. had previously shown

using an *in vitro* cleavage assay that it was possible to mutate the sequence of the stem loop of the direct repeat, as long as base-pairing was preserved (Zetsche et al., 2015a). Of particular interest, one of the mutation pairs they tested (mutating the 5th nucleotide from U to C and the 18th nucleotide from A to G) ablated the non-canonical termination sequence but showed no observable reduction in cleavage activity.

As such we proceeded to generate two new sets of arrays, one set which would utilise the original direct repeat and a second, using the same spacer sequences, but with mutated direct repeats, to remove the non-canonical termination sequence. 6 9-crRNA was designed for each set, where the most active single crRNA for Ascl1 (A2), HBB (H2) and IL1RN (I2) were incorporated at either the 1st, 5th or 9th position within the crRNA array, in all 6 possible combinations (AHI, AIH, HAI, HIA, IAH, IHA) (Figure 5.5A). This design would be expected to clearly show a stepwise reduction in activity for crRNAs at the 1st, 5th and 9th position, whilst also making it easier to observe when subsequent modifications successfully recover activity for crRNAs positioned towards the 3' of the array.

The remaining 6 positions within the array were occupied by non-targeting crRNAs generated as previously described. Finally, a positive control 3-crRNA array was made, using the original direct repeats, with A2 in position 1, H2 in position 2 and I2 in position 3. This 3-crRNA array was expected to enable robust transactivation for all three genes to be targeted. The arrays were constructed as previously described, by annealing oligo pairs (Table 5.3) and ligating them into a BpiI digested U6 vector

(Materials and Methods 2.1.1 and 2.1.2). Due to challenges when cloning the constructs, a decision was made to proceed with only 5 of the 6 crRNA constructs, AHI, AIH, HAI, HIA, and IAH. After sequence verification, the 6-crRNA constructs possessing the original direct repeats were transfected into HEK293, alongside dFnCas12a-VPR, with dFnCas12a-VPR only serving as a negative control. Three days post transfection, RNA was extracted, cDNA generated and qRT-PCR performed as previously described.

Table 5.3 - Oligos used for constructing 9-crRNA arrays with original direct repeat

AHI O array	**AIH O array**	**HAI O array**	**HIA O array**	**IAH O array**
Fn 1 A1 orig F	Fn 1 A1 orig F	Fn 1 H1 orig F	Fn 1 H1 orig F	Fn 1 I1 orig F
Fn 1 A1 orig R	Fn 1 A1 orig R	Fn 1 H1 orig R	Fn 1 H1 orig R	Fn 1 I1 orig R
Fn 2 A1 orig F	Fn 2 A1 orig F	Fn 2 H1 orig F	Fn 2 H1 orig F	Fn 2 I1 orig F
Fn 2 A1 orig R	Fn 2 A1 orig R	Fn 2 H1 orig R	Fn 2 H1 orig R	Fn 2 I1 orig R
Fn 3 1-2 orig F	Fn 3 1-2 orig F	Fn 3 1-2 orig F	Fn 3 1-2 orig F	Fn 3 1-2 orig F
Fn 3 1-2 orig R	Fn 3 1-2 orig R	Fn 3 1-2 orig R	Fn 3 1-2 orig R	Fn 3 1-2 orig R
Fn 4 2-3 orig F	Fn 4 2-3 orig F	Fn 4 2-3 orig F	Fn 4 2-3 orig F	Fn 4 2-3 orig F
Fn 4 2-3 orig R	Fn 4 2-3 orig R	Fn 4 2-3 orig R	Fn 4 2-3 orig R	Fn 4 2-3 orig R
Fn 5 H5 orig F	Fn 5 I5 orig F	Fn 5 A5 orig F	Fn 5 I5 orig F	Fn 5 A5 orig F
Fn 5 H5 orig R	Fn 5 I5 orig R	Fn 5 A5 orig R	Fn 5 I5 orig R	Fn 5 A5 orig R
Fn 6 H5 orig F	Fn 6 I5 orig F	Fn 6 A5 orig F	Fn 6 I5 orig F	Fn 6 A5 orig F
Fn 6 H5 orig R	Fn 6 I5 orig R	Fn 6 A5 orig R	Fn 6 I5 orig R	Fn 6 A5 orig R
Fn 7 4-5 orig F	Fn 7 4-5 orig F	Fn 7 4-5 orig F	Fn 7 4-5 orig F	Fn 7 4-5 orig F
Fn 7 4-5 orig R	Fn 7 4-5 orig R	Fn 7 4-5 orig R	Fn 7 4-5 orig R	Fn 7 4-5 orig R
Fn 8 5-6 orig F	Fn 8 5-6 orig F	Fn 8 5-6 orig F	Fn 8 5-6 orig F	Fn 8 5-6 orig F
Fn 8 5-6 orig R	Fn 8 5-6 orig R	Fn 8 5-6 orig R	Fn 8 5-6 orig R	Fn 8 5-6 orig R
Fn 9 I9 orig F	Fn 9 H9 orig F	Fn 9 I9 orig F	Fn 9 A9 orig F	Fn 9 H9 orig F
Fn 9 I9 orig R	Fn 9 H9 orig R	Fn 9 I9 orig R	Fn 9 A9 orig R	Fn 9 H9 orig R
Fn 10 I9 orig F	Fn 10 H9 orig F	Fn 10 I9 orig F	Fn 10 A9 orig F	Fn 10 H9 orig F
Fn 10 I9 orig R	Fn 10 H9 orig R	Fn 10 I9 orig R	Fn 10 A9 orig R	Fn 10 H9 orig R

(Refer to materials and methods section (Table 2.1) for sequence information.)

We screened these arrays measuring the relative abundance of IL1RN transcript and as when testing the 6-crRNA arrays, we were able to see robust transactivation for all crRNA arrays relative to the non-targeting negative control (Figure 5.5B). As observed with the 6-crRNA array we also saw a significant trend for diminished activity when the IL1RN crRNA was positioned closer to the 3' of the crRNA array (P = 0.0001).

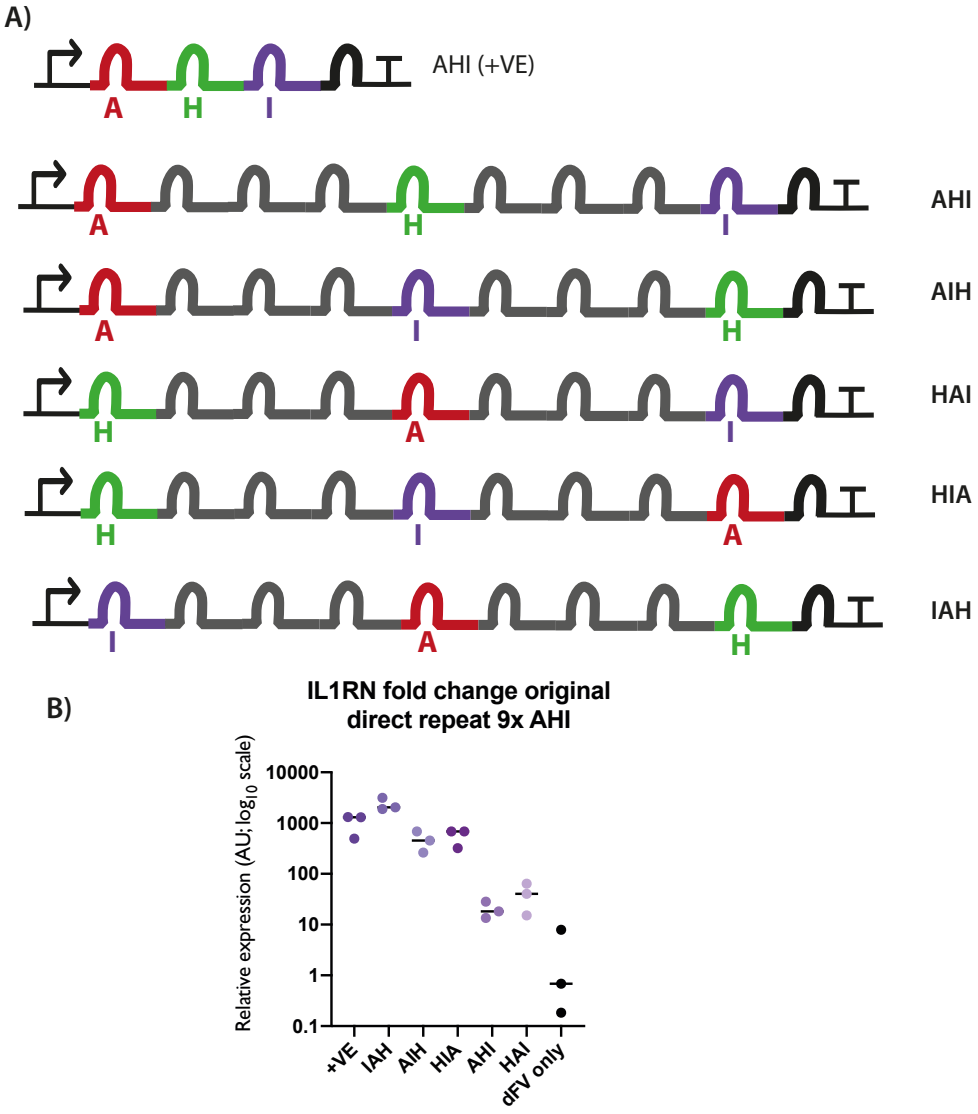

Figure 5-5 - Generating and screening 9x crRNA arrays [original Fn direct repeats]

A) Diagrammatic representation of the 9X crRNA arrays utilising the original Fn direct repeat sequences. B) Graph showing qRT-PCR data for the relative expression of IL1RN transcripts. The mean for the negative control (dFV NT) was used to normalise the remaining samples (n = 3).

We proceeded to clone crRNA arrays that were identical in design to the arrays just tested, with the exception that all direct repeats were mutated to ablate the non-canonical terminator sequence (Figure 5.6A), replacing the 5th nucleotide in the direct repeat with a 'C' and the 18th nucleotide in the direct repeat with a 'G'. These arrays were generated as previously described utilising annealed oligos which were ligated into a BpiI digested U6 vector plasmid. After the constructs were sequence verified, the relative abundance of IL1RN was screened for the 9-crRNA arrays possessing the mutated direct repeats. No significant increase in relative expression was observed for any of the crRNA arrays tested, with the exception of the positive control, which showed highly significant transactivation (Figure 5.6B).

Table 5.4 - Oligos used for constructing 9-crRNA arrays with original direct repeat

AHI M array	**AIH M array**	**HAI M array**	**HIA M array**	**IAH M array**	**IHA M array**
Fn 1 A1	Fn 1 A1	Fn 1 H1	Fn 1 H1	Fn 1 I1	Fn 1 I1
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 1 A1	Fn 1 A1	Fn 1 H1	Fn 1 H1	Fn 1 I1	Fn 1 I1
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R
Fn 2 A1	Fn 2 A1	Fn 2 H1	Fn 2 H1	Fn 2 I1	Fn 2 I1
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 2 A1	Fn 2 A1	Fn 2 H1	Fn 2 H1	Fn 2 I1	Fn 2 I1
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R
Fn 3 1-2	Fn 3 1-2	Fn 3 1-2	Fn 3 1-2	Fn 3 1-2	Fn 3 1-2
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 3 1-2	Fn 3 1-2	Fn 3 1-2	Fn 3 1-2	Fn 3 1-2	Fn 3 1-2
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R
Fn 4 2-3	Fn 4 2-3	Fn 4 2-3	Fn 4 2-3	Fn 4 2-3	Fn 4 2-3
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 4 2-3	Fn 4 2-3	Fn 4 2-3	Fn 4 2-3	Fn 4 2-3	Fn 4 2-3
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R
Fn 5 H5	Fn 5 I5	Fn 5 A5	Fn 5 I5	Fn 5 A5	Fn 5 H5
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 5 H5	Fn 5 I5	Fn 5 A5	Fn 5 I5	Fn 5 A5	Fn 5 H5
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R
Fn 6 H5	Fn 6 I5	Fn 6 A5	Fn 6 I5	Fn 6 A5	Fn 6 H5
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 6 H5	Fn 6 I5	Fn 6 A5	Fn 6 I5	Fn 6 A5	Fn 6 H5
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R
Fn 7 4-5	Fn 7 4-5	Fn 7 4-5	Fn 7 4-5	Fn 7 4-5	Fn 7 4-5
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 7 4-5	Fn 7 4-5	Fn 7 4-5	Fn 7 4-5	Fn 7 4-5	Fn 7 4-5
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R
Fn 8 5-6	Fn 8 5-6	Fn 8 5-6	Fn 8 5-6	Fn 8 5-6	Fn 8 5-6
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 8 5-6	Fn 8 5-6	Fn 8 5-6	Fn 8 5-6	Fn 8 5-6	Fn 8 5-6
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R
Fn 9 I9	Fn 9 H9	Fn 9 I9	Fn 9 A9	Fn 9 H9	Fn 9 A9
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 9 I9	Fn 9 H9	Fn 9 I9	Fn 9 A9	Fn 9 H9	Fn 9 A9
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R
Fn 10 I9	Fn 10 H9	Fn 10 I9	Fn 10 A9	Fn 10 H9	Fn 10 A9
Mut F	Mut F	Mut F	Mut F	Mut F	Mut F
Fn 10 I9	Fn 10 H9	Fn 10 I9	Fn 10 A9	Fn 10 H9	Fn 10 A9
Mut R	Mut R	Mut R	Mut R	Mut R	Mut R

(Refer to materials and methods section (Table 2.1) for sequence information.)

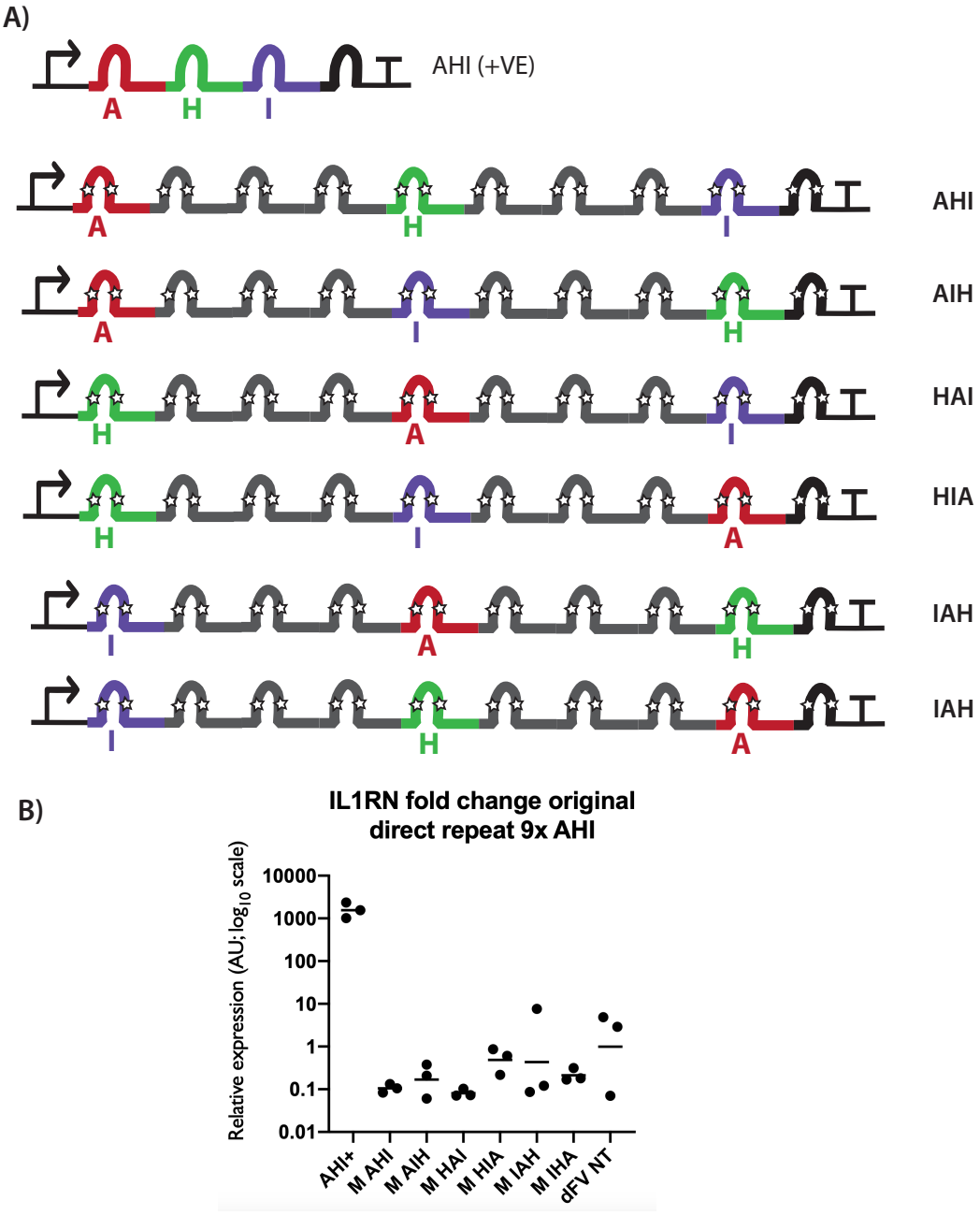

Figure 5-6 - Generating and screening 9x crRNA arrays [mutated Fn direct repeats]

A) Diagrammatic representation of the 9X crRNA arrays utilising mutated Fn direct repeat sequences. Stars signify reciprocal mutation in the stem loop of the 5th and 18th nucleotides. B) Graph showing qRT-PCR data for the relative expression of IL1RN transcripts. The mean for the negative control (dFV NT) was used to normalise the remaining samples (n = 2).

5.7 Generating split dFnCas12a-VPR

One of the key future applications of synthetic transcription factors is for gene therapies. Whilst it has been shown that both Cas9 and Cas12a based synthetic transcription factors can be delivered to human cells outside of the body, when considering gene therapies *in vivo*, delivery becomes more challenging. Adeno-associated viruses (AAV) offer one of the most attractive delivery vehicles, showing exceptionally low immunogenicity. However, AAV viruses also provide highly restrictive size constraints, being able to package only 4.4Kb of genetic material. This means delivery of the dFnCas12a-VPR constructs at their current size (~5.6Kb) would be impossible. One strategy to address this problem stems from work with Cas9, where researchers have shown it is possible to split Cas9 into 2 pieces (Zetsche et al., 2015c) which still show cleavage activity. Further work also showed it was possible to split a dCas9-VP64 synthetic transcription factor into two parts, and by fusing an FRB domain to the 1st half and FKBP domain to the 2nd half they were able to see robust transactivation when the two halves were expressed from different plasmids (Zetsche et al., 2015c). As such we aimed to test whether it was possible to see any transactivation when split dFnCas12a-VPR halves were expressed from different plasmids. 4 split variants of dFnCas12a-VPR were designed (Figure 5.7A), with the split sites being rationally selected. The split sites were chosen to be close to the middle of the coding sequence, whilst lying within an exposed loop of the protein (shown in Figure 5.7B) as previous work has shown that successful split sites across all proteins tested lay within exposed loops (Dagliyan et al., 2018).

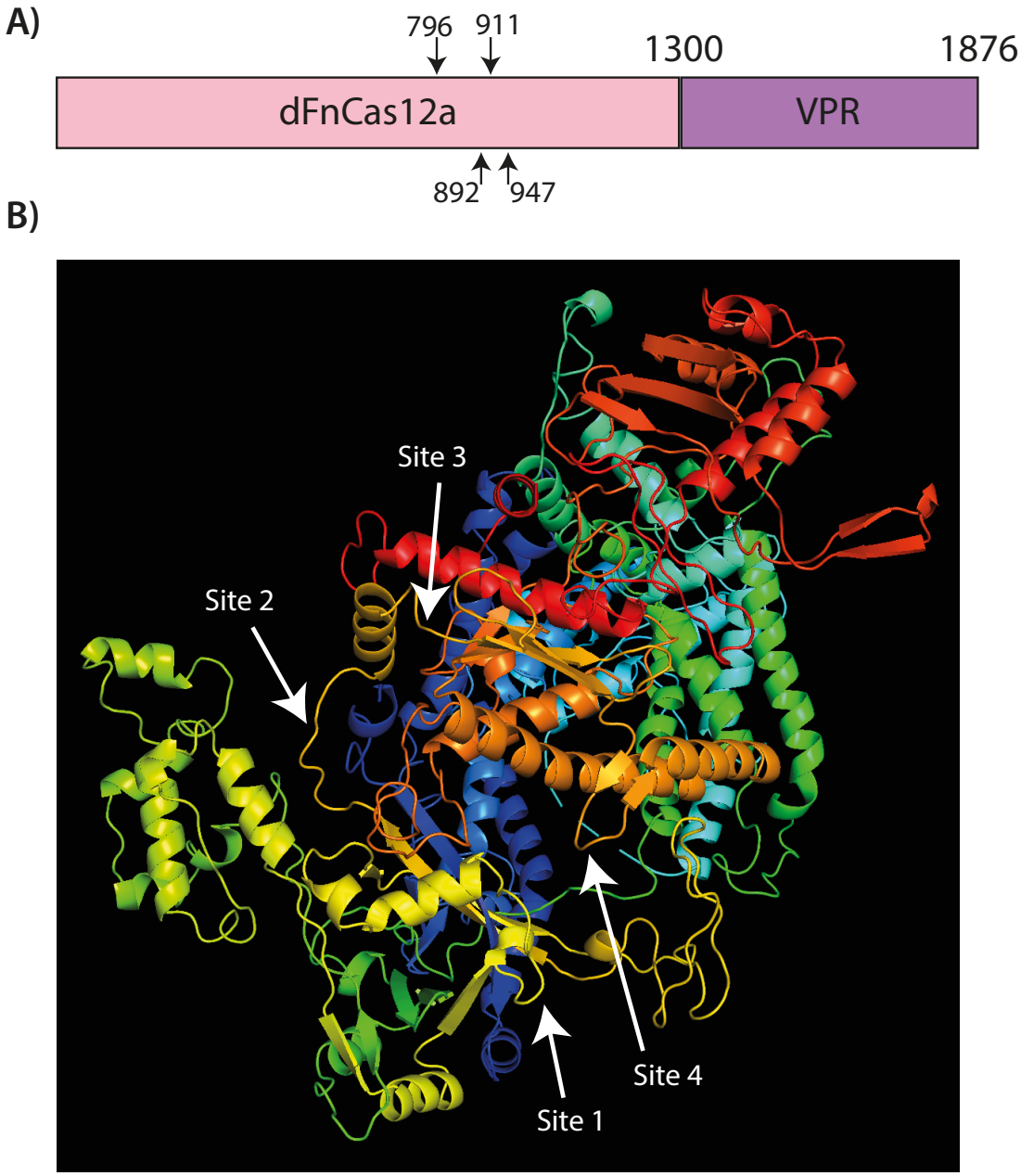

Figure 5-7 - Design of split dFnCas12a-VPR variants

A) Diagrammatic representation of the positions of the 4 different split sites with full length dFnCas12a-VPR. The numbers highlight the amino acid position within the coding sequence. B) Image showing the protein visualisation of wild-type FnCas12a. The arrows highlight the positions of the 4 split sites within the protein structure.

As a first step, to facilitate the library generation. We designed and generated an acceptor plasmid, where two Esp3I restriction sites were present immediately downstream of a CmV promoter and immediately upstream of a bGH terminator. This was achieved by digesting the FnCas12a original plasmid using Kpn1 and Xho1 to remove the coding sequence between the CmV promoter and bGH terminator. The plasmid was then repaired with annealed oligos that incorporated two Esp3I restriction sites, to facilitate subsequent cloning of split dFnCas12a-VPR coding sequences (Figure 5.8A). This enabled the generation of all subsequent 'halves' by PCR amplification of the corresponding region of the full length dFnCas12a-VPR, with the primers designed to flank the amplified sequences with BpiI sites and include a stop codon at the ends of the amplified coding sequence for the first halves and a start codon before the amplified coding sequence for the second halves. PCR amplification from the dFnCas12a-VPR plasmid was performed and for all 8 fragments. The PCR reactions were subsequently column purified and BpiI digested before a second PCR purification was carried out to remove the unwanted digested fragments at either ends of the PCR products. These digested amplicons were subsequently ligated into the Esp3I digested 'CmV Esp3I Ter' acceptor vector (Materials and Methods 2.1.9 and 2.1.2) (Figure 5.8B). After sequence verification of the 8 constructs, they were transfected into HEK293 cells alongside the full length dFnCas12a-VPR.

The screen was set up such that either the first half plasmid, the second half plasmid or both halves were transfected. For the positive control (full length dFnCas12a-VPR) and all test conditions the single crRNA 'H1' was also delivered.

The full length dFnCas12a-VPR without a targeting crRNA served as a negative control. Three days post transfection, RNA was extracted, cDNA generated and qRT-PCR performed as previously described (Materials and Methods 2.3.2).

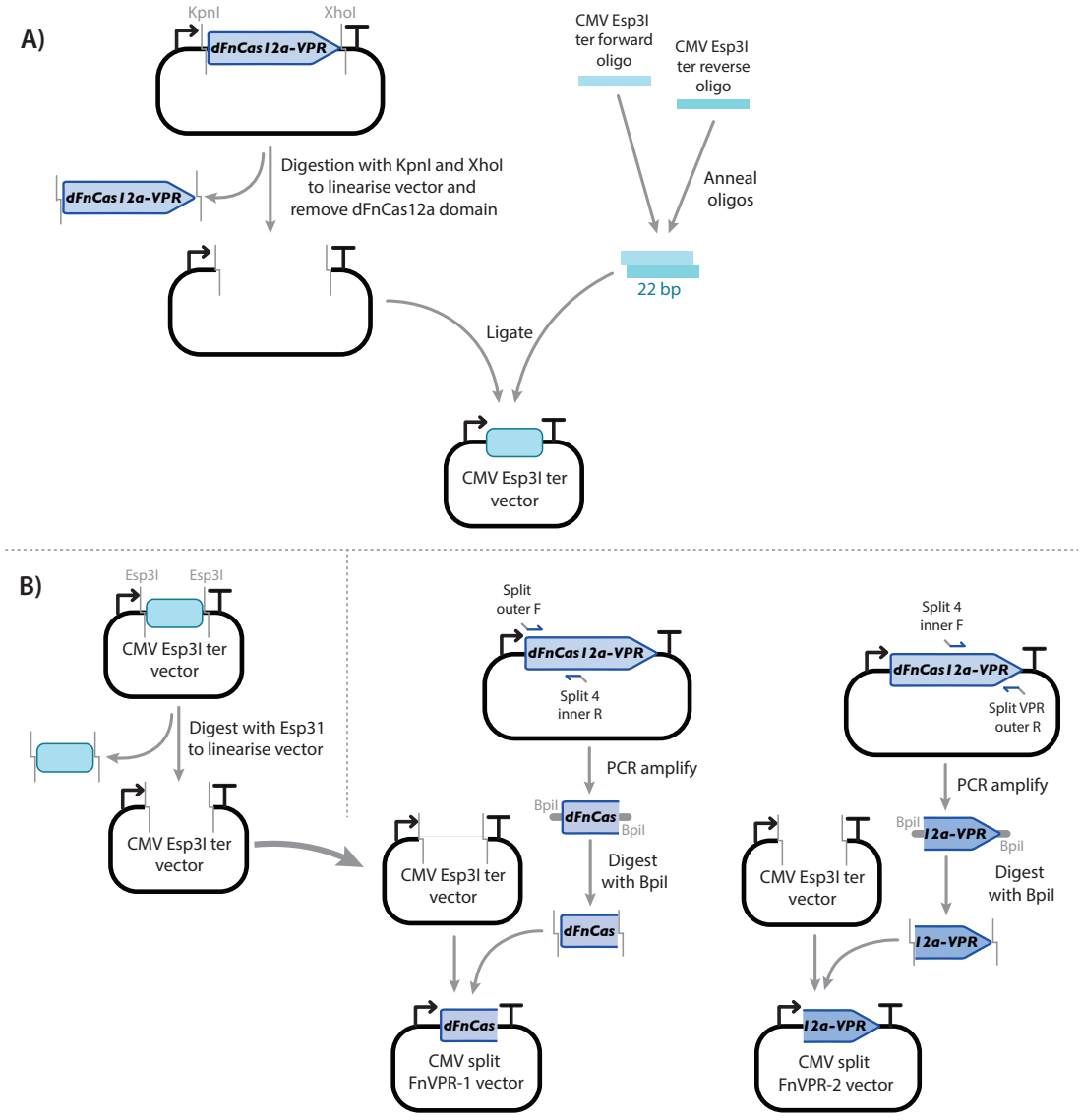

Figure 5-8 - Cloning of split dFnCas12a-VPR variants

A) Diagrammatic representation of cloning strategy for generating the CmV Esp3I Ter vector for receiving split dFnCas12a-VPR fragments B) Diagrammatic representation of cloning strategy for generating the split dFnCas12a-VPR plasmids. A) Graphs showing the qRT-PCR data for the full length dFnCas12a-VPR delivered with the H1 crRNA serving a positive control and the full length dFnCas12a-VPR without a targeting crRNA served as a normalising negative control (n = 2). The first and second 'halves' are delivered individually or together.

An ordinary one-way ANOVA test was performed, with a subsequent Tukey's multiple comparisons test showing that with the exception of the full length dFnCas12a-VPR positive control, only the condition with both split halves present for the split at position 2 and the split at position 4 showed significant transactivation above the dFnCas12a-VPR only negative control (P = 0.04 and P = 0.0003 respectively). Furthermore, when comparing the conditions where both halves were delivered compared to either the first half or the second half, only the split at position 4 showed significantly greater activation than if either half was independently delivered (P = 0.03 and P = 0.03 respectively). These results are visually represented in Figure 5.9.

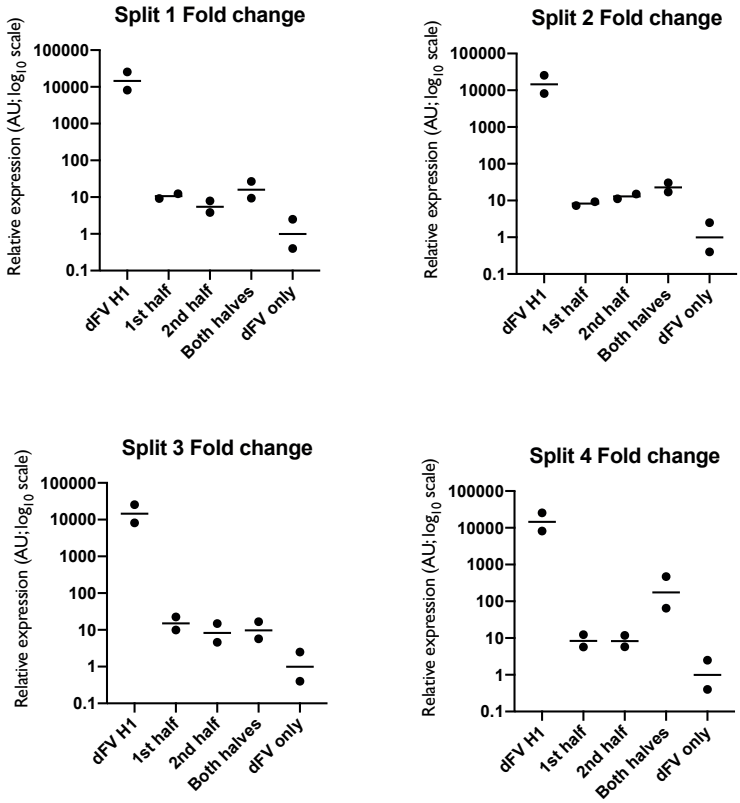

Figure 5-9 - Screening split dFnCas12a-VPR variants

qRT-PCR data for the full length dFnCas12a-VPR delivered with the H1 crRNA serving a positive control and the full length dFnCas12a-VPR without a targeting crRNA served as a normalising negative control (n = 2). The first and second 'halves' are delivered individually or together.

5.8 Generating synthetic chromatin editors

CRISPR based synthetic chromatin editors, as previously discussed in chapter 1, are chromatin modifying effector domains, recruited by dCas9 to a target locus to enable the laying or removal of a specific epigenetic mark. This enables a more directly related analysis of causality when compared to chemical approaches for epigenetic modification, where global changes are observed. In the following work we looked to generate novel synthetic chromatin editors, aiming to lay a trimethylation mark at histone 3, lysine 36 (H3K36me3). This mark has typically been associated with heterochromatin and repression of promoters, however of interest, work by Luco et al. demonstrated that through overexpression of SETD2, they were able to observe not only increased laying of the H3K36me3 mark, but furthermore significantly altered exon inclusion rates for polypyrimidine tract–binding protein dependent exons (Luco et al., 2010). Suggesting there may be a direct causality between the laying of the H3K36me3 and alternative splicing. To explore this relationship, we sought to generate a synthetic chromatin editor that was able to lay the H3K36me3 mark at specific loci. To achieve this, the catalytic core of SETD2 (Yang et al., 2016) was amplified from a GFP-SETD2 plasmid (Carvalho et al., 2014) using SETD2 F + SETD2 R (Table 2.1). The PCR was subsequently ran on a 1% agarose gel and the amplicon was gel extracted. Next the dcas9-VPR plasmid (addgene #63798) (Table 2.2) was digested with BsrGI to release the VPR domain. After incubation the digestion was run on an agarose gel and the plasmid backbone was gel extracted and purified. Finally the purified backbone and

the SETD2 core amplicon were assembled using Gibson assembly (Figure 5.10A) (Materials and Methods 2.1.13).

We also designed and generated a dCas9-SETD1A synthetic chromatin editor, as SETD1A has been shown to be responsible for mono, bi and tri-methylation of histone 3, lysine 4 (H3K4) (Lee and Skalnik, 2005). A correlation between the H3K4me3 and alternative splicing has been observed, with suppression of H3K4me3 through ASH2L knockdown impacting splicing patterns and suppression of splicing attenuating H3K4me3 (Davie et al., 2015). A putative catalytic core of SETD1A was chosen, including the set domain as well as the adjacent CFP1 interacting domain, shown to be important for SETD1A activity (Lee and Skalnik, 2008).

The dCas9-SETD1A was generated by ordering a gblock (IDT) encoding the SETD1A core domain. This was subsequently PCR amplified using SETD1A F + SETD1A R (Table 2.1). The dCas9-VPR plasmid was digested with BsrGI to release the VPR domain. After incubation the digestion was run on an agarose gel and the plasmid backbone was gel extracted and purified. Gibson assembly was performed, using the linearised plasmid backbone and the SETD1A core amplicon to generate the dCas9-SETD1A plasmid (Figure 5.10A).

After sequence verification, dCas9-SETD2, dCas9-SETD1A and dCas9-VPR were transfected into HEK293 cells. The transiently transfected cells were incubated for 2 days at 37°C and 5% CO_2, before being washed with PBS and resuspended in SDS-PAGE sample buffer for Western Blot analysis using an anti Cas9 polyclonal

antibody (Clontech GuideIT, 632607) (1:3000) (Materials and Methods 2.1.14). The proteins could then be visualised alongside the Beta actin loading control, after secondary incubation with anti-rabbit HRP antibody and HRP conjugated anti-beta-actin (SIGMA, BA3R MA5-15739-HRP) (1:10,000) (Figure 5.10B).

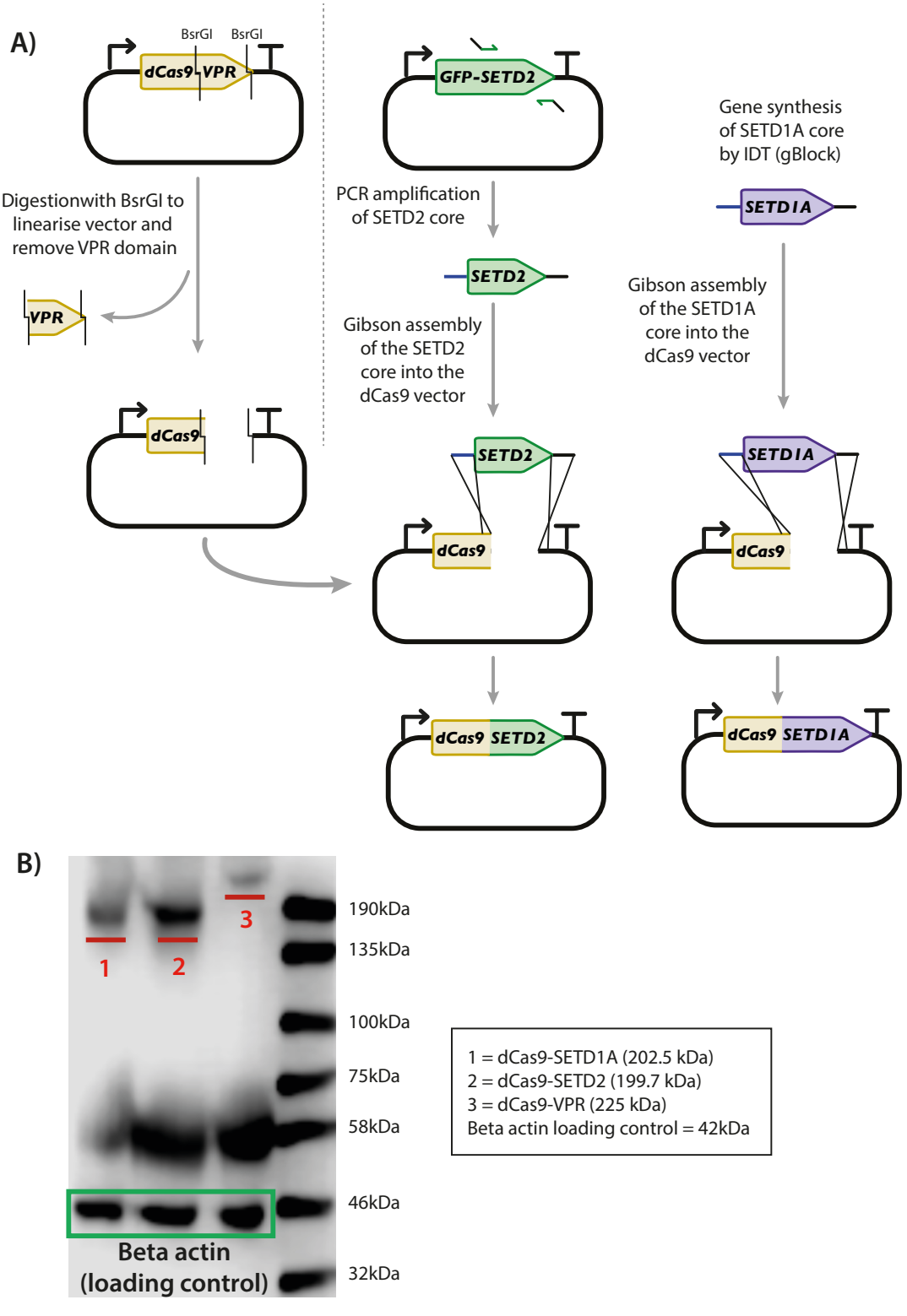

Figure 5-10 - Cloning and screening for expression of putative synthetic chromatin editors

A) Diagrammatic representation of cloning strategy for generating the putative synthetic transcription factors dCas9-SETD2 and dCas9-SETD1A B) Western blot showing the expression of dCas9-SETD1A, dCas9-SETD2 and dCas9-VPR in HEK293 cells.

After showing that the putative synthetic transcription factors successfully expressed in mammalian cells, we proceeded to establish a ChIP-qPCR protocol (described in greater detail in chapter 2), to enable identification of differences in deposition of a mark at the target loci. In the first stage of this assay, chromatin is crosslinked using formaldehyde, to preserve the interaction between histones and target DNA over the subsequent steps. After crosslinking, the cells are sonicated, to produce chromatin fragments. At this stage a quality control step is performed, with the ideal shearing pattern to lie between 200bp and 500bp (Figure 5.11A), to provide high resolution without significant disruption of nucleosomes. If the chromatin is of sufficient quality and size, then immunoprecipitation is performed to enable preferential enrichment of loci crosslinked with histones possessing the mark of interest. Finally, the enriched chromatin can be purified alongside un-enriched chromatin and qPCR can be performed to screen for changes of abundance at a loci of interest or preferential enrichment of two or more targets. With the help of a colleague at the Buonomo lab, we screened for enriched histone 3 lysine 27 acetylation at a positive control locus (TSIX), relative to a negative control loci (X5) (previously validated by our collaborator) (Table 2.3). We were subsequently able to observe a highly significant enrichment for the TSIX locus ($P = 0.002$) (Figure 5.11B), establishing the ChIP qPCR protocol as a viable screening technique for the exploration of enrichment of specific chromatin marks.

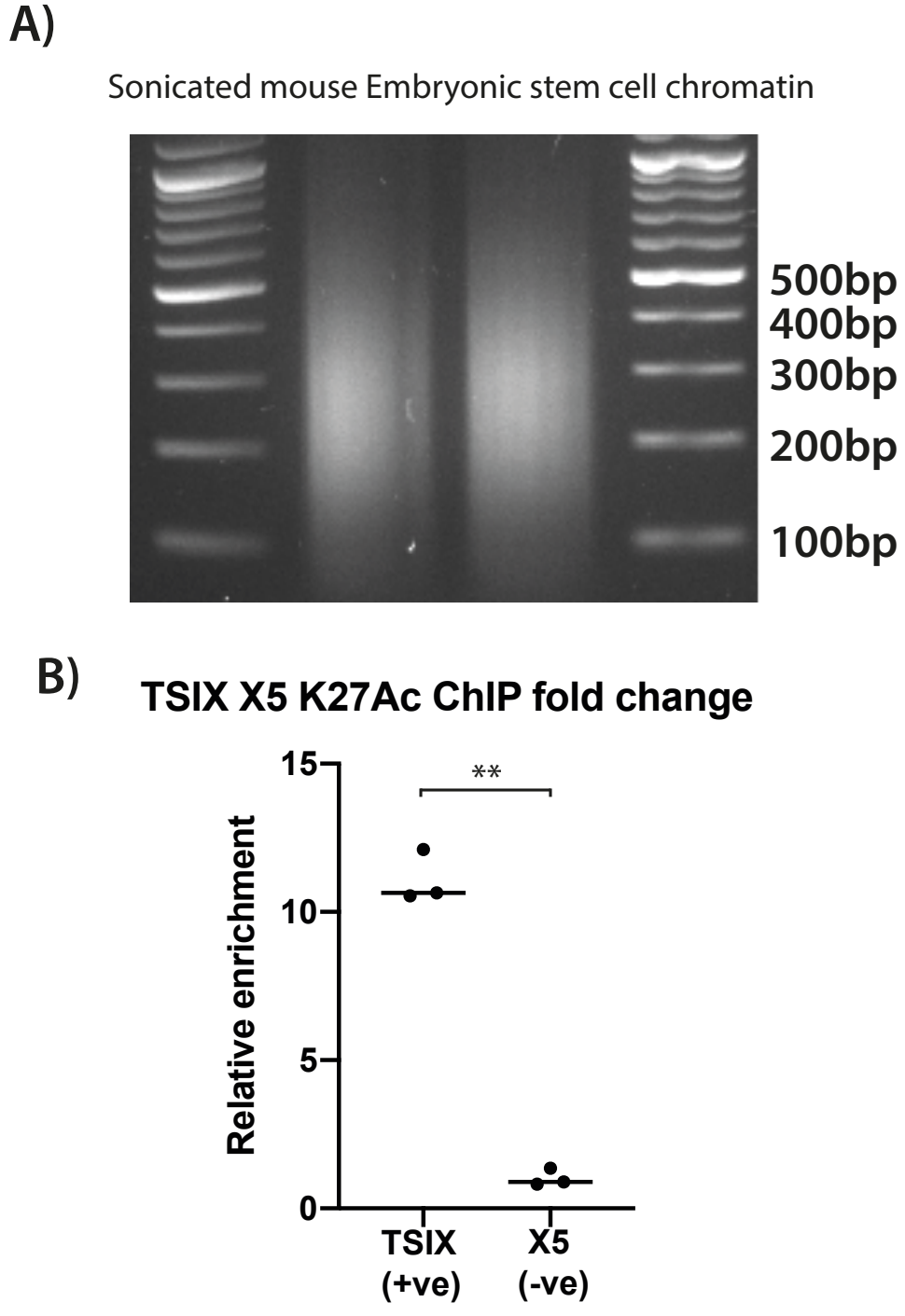

Figure 5-11 - Establishing ChIP-qPCR protocol

A) Sonicated mouse embryonic stem cell chromatin ran on a 1% agarose gel B) Graph showing qRT-PCR results, for the relative enrichment of target loci after H3K27ac enrichment of chromatin (n = 3). The ratio of the abundance of each loci in the enriched chromatin fraction compared to an unenriched control was calculated, before normalising the results to the negative control (X5) set to 1.

5.9 Conclusion and Discussion

In the preceding chapter we have explored synergy, observing significant synergy occurring for two of the four crRNA arrays tested, but not when delivering the crRNA pairs expressed from separate plasmids. This discrepancy may in part be due to the fact that the pooled crRNAs were delivered on separate plasmids, whereas each array was expressed from a single plasmid. As transfection efficiencies are below 100% a lower percentage of cells can expect to receive both crRNA plasmids at the same time (pooled condition) compared to a single plasmid and as such any synergistic impact cause by co-delivery of separate plasmids will be reduced.

We also explore the capabilities of crRNA arrays to enable transactivation for dFnCas12a-VPR. We not only see robust activity for all arrays tested but can also observe multiplexed transactivation of three different genes from a single 6-crRNA array. We do however observe a significant drop in activity when crRNAs are positioned to the 3' of longer crRNA arrays. We go on to explore a possible cause, the presence of a non-canonical pol3 termination sequence within the direct repeats of the crRNA arrays. However due to the lack of observable activity seen when the direct repeat is mutated (based upon results from Zetsche et al. who showed that *in vitro* cleavage was preserved with this specific reciprocal mutation), we chose not to pursue this further. In the future it would be interesting to explore if the same position dependent reduction in activity is observed when the crRNA array is expressed from a pol2 promoter. Whilst Campa et al. have recently shown

that long 20-crRNA arrays can be expressed from a pol2 transcript, they did not explore or otherwise observe position dependent activity within their arrays (Campa et al., 2019).

We also provide evidence that split variants of dFnCas12a-VPR are able to induce transactivation when co-expressed in the same cells alongside a targeting crRNA. This capability opens up a number of applications for dFnCas12a-VPR, allowing delivery using AAV viruses (preferred for gene therapy due to low immunogenicity) and providing a substrate for layered logic gates, as both halves must be expressed to observe activity.

Of note more work will need to be carried out to see if the same split 'halves' show significant transactivation when targeting alternative genes. Furthermore, the level of transactivation is severely diminished for the split variants relative to the full length dFnCas12a-VPR a key limitation that could be addressed either through the incorporation of dimerisation domains (Zetsche et al., 2015c) or inteins. Both the use of inteins and dimerisation domains have shown the capacity to recover activity close the full length proteins for split dCas9-VP64 and wild-type Cas9 respectively.

Finally, whilst we generated and showed expression for putative synthetic chromatin editors and established a ChIP qPCR pipeline, we have yet to screen the putative synthetic chromatin editors for activity. In future work, these constructs would be screened using ChIP-qPCR and if active, the same pipeline can be used to test the effector domains tethered to the dFnCas12a scaffold.

Chapter 6 - Discussion

In this body of work we set out with the aim of generating Cas12a derived synthetic transcription factors, with the subsequent aims of characterising and working towards applications. In chapter 3, we confirm work that was published during the project by Tak et al. (Tak et al., 2017), showing that dAsCas12a can be adapted as a synthetic transcription factor. However, we provided a further two novel findings, showing that transactivation for dAsCas12a could be achieved when the transactivation domain was tethered to the N-terminus of the protein. The fact that dAsCas12a is still able to bind and VPR can still transactivate when fused to the N-terminus suggests that this may be the case for other effector domains. This information can inform design considerations, when there are concerns about misfolding of the fusion protein or inactivation of the effector domain. Whilst dCas9-VPR and dCas12a-VPR are screened side by side, targeting the same Luciferase vectors, it is challenging to draw meaningful comparison of their relative activity. There are a large number of factors known to be involved in gRNA activity for synthetic transcription factors; chromatin state, positioning relative to TSS, position of effector domain relative to the CRISPR scaffold, PAM sequence selected etc. (Gilbert et al., 2014). As such it becomes very challenging to perform a meaningful comparative screen of Cas9 derived synthetic transcription factors and Cas12a derived synthetic transcription factors. Meaningful analysis will only be possible when employing a large-scale screen, so that a number of factors can be

simultaneously explored, with sufficient data points to derive significant conclusions.

In chapter 4 we also demonstrated the novel finding that FnCas12a can be adapted to produce a synthetic transcription factor. This variant was of special interest for two reasons, firstly, with the exception of the original characterisation by Zetsche et al. (Zetsche et al., 2015a), the Fn variant has been understudied due to original work by Zetsche et al. suggesting it was not able to cleave the mammalian genome. Secondly and of critical importance, the Fn variant was characterised to have a shorter PAM sequence 'TTN' than the As and Lb variants ('TTTN'), that have been the primary focus of subsequent research. As previously described, this feature enables denser targeting of genomic loci, providing approximately 4 times as many potential binding sites. This in turn provides more opportunities and flexibility when considering applications where exact positioning is important (such as transactivation) or when considering targeting multiple different targets simultaneously, where the chances of finding an active crRNA for all loci will decrease as the targets being screened increases.

In chapter 4 we also screened the C-terminally tagged dAsCas12a-VPR, dFnCas12a-VPR and dLbCas12a-VPR variants side by side and observed that all 3 variants showed significant transactivation. We also tested the activity of the 3 different variants when delivered with their own crRNAs or crRNAs with direct repeat sequences derived from the remaining two variants to assess orthogonality. Whilst we observed high fidelity (minimal activity with non-native crRNAs) for dFnCas12a-

VPR, we did not observe high fidelity for dAsCas12a-VPR, which induced significant transactivation when delivered with the Fn crRNA, nor dLbCas12a-VPR, as activity was just below significance (P = 0.051).

Orthogonality provides the capacity to enact independent programs within a cell, for example allowing the capacity to transactivate one group of genes and simultaneously repress another group of genes. Possessing this level of control is vital when seeking to generate predictable synthetic circuits and is essential for applications such as the laying of epigenetic marks. Orthogonality between different chromatin editors means that you can be confident only the desired chromatin mark is being laid at desired loci when utilising multiple chromatin editors within the same cell.

In chapter 4 we go on the show that dFnCas12a-VPR can transactivate 3 different target genes using a single crRNA per target. Of particular note, we were able to see significant transactivation when targeting the PAM sequences; 'TTTA' (H4 crRNA), 'TTTC' (I2 crRNA) and 'TTTG' (A2, H1, H2), similar to observations seen for AsCas12a and LbCas12a (Kim et al., 2017).

We go on to observe that improved activation of a target locus can be achieved when delivering multiple crRNA targeting a single gene, similar to results observed with dLbCas12a-VPR (Tak et al., 2017). At the conclusion of chapter 4 we demonstrate that dFnCas12a-VPR can enable transactivation when crRNA arrays are delivered, showcasing one of the main strengths of the Cas12a system relative

to Cas9, through the ability to process its own crRNA array for subsequent targeting.

In chapter 5 we focus on characterisation. First addressing the question of synergy, defined as the capacity to induce greater transactivation than would be expected by adding together the transactivation observed for the individual crRNAs. Collaborating with a statistician, we sought to provide a realistic estimation of not only the mean of an additive distribution (simply calculated by adding the mean fold increases of the individual crRNAs) but also the variance of the hypothetical distribution. This contrasts to previous approaches where the hypothetical additive variance was not derived from the constituent activity of the individual crRNAs, but by assuming that the variance would be the same as the experimental combined condition (Tak et al., 2017). When assessing co-transfected crRNA pairs, we did not observe synergy, however when the crRNA pairs were expressed from a single array we did observe synergistic transactivation for two of the 4 arrays tested.

We go on to further explore the capabilities of crRNA arrays, testing multiplexing, where we can observe significant transactivation for three different genes targeted from the same array. This is consistent with results observed for dLbCas12a-VPR (Tak et al., 2017). We also saw that significant transactivation could be achieved when using 9-crRNA arrays. However, for all 6-crRNA and 9-crRNA arrays tested, we could observe a clear drop in activity for crRNAs positioned towards the 3' of the array. This is a very important consideration when utilising Cas12a derived synthetic transcription factors, which can be considered during experimental design. One

example would be placing less active crRNAs for genes that show weaker transactivation at the 5' of an array and more active crRNAs for genes that show stronger transactivation at the 3' of the array it will be possible to up-regulate all target genes. There are a number of possible reasons as to why this consistent drop in activity is observed. One possibility is that this is a problem unique to polymerase 3 driven promoters, with premature termination being seen when the polymerase recognises a weak non-canonical termination sequence in the direct repeats of the crRNA array. We began to explore this phenomenon, by mutating the direct repeats within a longer array, mutating a non-canonical termination sequence identified by Orioli et al. (Orioli et al., 2011) 'TTTCT'. The mutation was utilised 2 nucleotide substitutions in the direct repeat identified by Zetsche and colleagues (Zetsche et al., 2015a) to preserve cleavage activity of wild-type FnCas12a *in vitro*. Due to the complete loss of activity within the 9-crRNA array upon mutation, we decided not to further pursue this approach.

Finally, we looked into generating split variants of dFnCas12a-VPR. This is especially important when considering applications of the synthetic transcription factor. In terms of synthetic biology research, by splitting the synthetic transcription factor into two parts and having the two proteins expressed by different promoters, it becomes possible to perform a number of logic operations. This can be achieved by controlling expression of either half, through presence/absence, inversions using recombinases or expressing from independent chemically inducible promoters.

Clinical trials for the use of wild-type Cas9 are already underway and for a subset of debilitating diseases this offers a potential solution to improve a patient's condition. However, the induction of a double-stranded break at a target locus is quite a drastic solution, exerting stress/toxicity on cells and invariably elevating the risk of cancer through increased chromosomal re-arrangements. As such approaches that can manipulate the transcriptome without inducing a double strand break may offer broader therapeutic value. For example in cases where phenotypes can be improved by increasing expression of a gene that can partially or fully compensate for a non-functional gene (Perumbeti et al., 2009). However, CRISPR based synthetic transcription factors are often too large to package into the preferred viral delivery vector adeno-associated virus (AAV), which offers an ideal delivery vector due to the minimal immune response elicited. A key limitation however is the fact that AAV viruses are only able to package up to a maximum of 4.4Kb, with decreased packaging efficiency observed when approaching this upper limit. This presents a key challenge for SpCas9 derived synthetic transcription factors as well as the Cas12a derived synthetic transcription factors characterised in the preceding work, with dFnCas12a-VPR being encoded by 5.6Kb of DNA.

Whilst work with SaCas9-VP64 (Matharu et al., 2019) have demonstrated that by using a shorter CRISPR protein it is possible to see transactivation and improvements in mouse pathology *in vivo*, the group still only tested delivery of one gRNA at a time and delivered the gRNA using a second AAV virus.

Due to the fact that the dFnCas12a-VPR is 5.6Kb, we explored the generation of a split variant, where two halves of the full-length protein could be expressed as two different transcriptional units. This strategy was based upon work with the Cas9 system, previously described in Chapter 1 and Chapter 5, multiple groups were able to show that wild-type Cas9 could be split into two or even three pieces and still enable cleavage. In particular work by Zetsche et al. demonstrated that dCas9-VP64 could be split into two fragments (Zetsche et al., 2015c). Building upon this work, we demonstrate initial evidence that dFnCas12a-VPR can be split and expressed on different plasmids. Whilst more work will be carried out to confirm this is observed when targeting different genes, these initial results open up the potential to deliver a split dFnCas12a-VPR *in vivo* using AAV viruses. Recent work by Campa et al. showed that crRNA arrays can be expressed and processed from the same pol2 transcript as AsCas12a (Campa et al., 2019), incorporating a triplex secondary structure between the AsCas12a stop site and the crRNA array to prevent transcript degradation. By utilising a triplex between the split dFnCas12a-VPR 'halves' driven by a CmV minimal promoter, there would remain more than 1Kb of packaging space for the crRNA array, which represents over 25 targeting crRNAs. This would provide the flexibility to target multiple crRNAs to individual promoters to increase transactivation.

Ongoing work will also look to explore whether improvements in transactivation for split dFnCas12a-VPR can be achieved by incorporating split inteins (protein motifs that enable trans-splicing of two different polypeptides, forming a single protein) into the two halves. As the NpuN and SspC split intein pair are only 312 and 113 bp

respectively (Stevens et al., 2016) this would still enable the system to be utilised for AAV delivery *in vivo*.

Finally, one of the most complex layers of regulation within the cell is the epigenome, with chemical alterations of the chromatin enabling semi-heritable, often long-lasting alterations to nuclear architecture and transcriptional programs that can reshape the phenotype of a cell and its progenitors.

Of note it can be exceptionally challenging to decouple transcriptional manipulation and epigenetic manipulation. The most obvious causal relationships are caused by the fact that increasing transcription from a locus can lead to opening of a locus, reducing nucleosome occupancy and hence the association of histone associated chromatin marks. It is also important to remember that the polymerase complex responsible for reverse transcribing pol2 genes, is able to not only read epigenetic marks as it processes, but can also lay these marks. As a result, by increasing the polymerase density at a locus a number of confounding effects can be expected to occur as speed and durations of interactions are all altered. Having said this, we already know that epigenetic modifications are very important in regulation, and as with many features of synthetic biology, there are some approaches for manipulation that can offer relatively clear cut phenotypes. For example, by recruiting a histone acetyl-transferase to an inactive distal enhancer, it is possible to acetylate the nearby histones and induce a euchromatic state (Hilton et al., 2015), effectively re-activating the enhancer and enable subsequent activation of downstream genes and genetic programs. As the prices and technical challenges

associated with epigenetic analysis decrease, it is our expectation that there will be increasing interest in diverse tools to generate specific, targeted modifications. Further we believe that with improved understanding and more powerful predictive software, there will be a demand for platforms, such as dFnCas12a, that are able to densely target diverse loci with relative ease.

It is the author's hope that through the generation, characterisation and efforts towards applications of dFnCas12a synthetic transcription factors, we have aided in expanding the tools available for cell biologists to explore fundamental questions. Further we hope these tools can help accelerate the acquisition of knowledge around the manipulation of transcriptional networks, providing the knowledgebase to build towards clinical applications utilising this tool or comparable tools.

Chapter 7 References

Abugessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., Severin, J., Carninci, P., Kawaji, H., Kasukawa, T., 2017. FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. Sci. Data 4. https://doi.org/10.1038/sdata.2017.107

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science 315, 1709–1712. https://doi.org/10.1126/science.1138140

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A., Bonas, U., 2009. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509–1512. https://doi.org/10.1126/science.1178811

Campa, C.C., Weisbach, N.R., Santinha, A.J., Incarnato, D., Platt, R.J., 2019. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. Nat. Methods 16, 887–893. https://doi.org/10.1038/s41592-019-0508-6

Carvalho, S., Vítor, A.C., Sridhara, S.C., Martins, F.B., Raposo, A.C., Desterro, J.M., Ferreira, J., de Almeida, S.F., 2014. SETD2 is required for DNA double-strand break repair and activation of the p53-mediated checkpoint. eLife 3, e02482. https://doi.org/10.7554/eLife.02482

Chapdelaine, P., Coulombe, Z., Chikh, A., Gérard, C., Tremblay, J.P., 2013. A Potential New Therapeutic Approach for Friedreich Ataxia: Induction of Frataxin Expression With TALE Proteins. Mol. Ther. Nucleic Acids 2, e119. https://doi.org/10.1038/mtna.2013.41

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., Iyer, E.P.R., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., Ter-Ovanesyan, D., Braff, J.L., Davidsohn, N., Housden, B.E., Perrimon, N., Weiss, R., Aach, J., Collins, J.J., Church, G.M., 2015. Highly efficient Cas9-mediated transcriptional programming. Nat. Methods 12, 326–328. https://doi.org/10.1038/nmeth.3312

Cole, P.A., 2008. Chemical probes for histone-modifying enzymes. Nat. Chem. Biol. 4, 590–597. https://doi.org/10.1038/nchembio.111

Costello, J., Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., 2013. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins.

Dagliyan, O., Krokhotin, A., Ozkan-Dagliyan, I., Deiters, A., Der, C.J., Hahn, K.M., Dokholyan, N.V., 2018. Computational design of chemogenetic and optogenetic split proteins. Nat. Commun. 9, 1–8. https://doi.org/10.1038/s41467-018-06531-4

Davie, J.R., Xu, W., Delcuve, G.P., 2015. Histone H3K4 trimethylation: dynamic interplay with pre-mRNA splicing. Biochem. Cell Biol. 94, 1–11. https://doi.org/10.1139/bcb-2015-0065

Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.-K., Shi, Y., Yan, N., 2012. Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. Science 335, 720–723. https://doi.org/10.1126/science.1215670

Deuschle, U., Meyer, W.K., Thiesen, H.J., 1995. Tetracycline-reversible silencing of eukaryotic promoters. Mol. Cell. Biol. 15, 1907–1914. https://doi.org/10.1128/MCB.15.4.1907

Dominguez-Monedero, A., Davies, J.A., 2018. Tamoxifen- and Mifepristone-Inducible Versions of CRISPR Effectors, Cas9 and Cpf1. ACS Synth. Biol. 7, 2160–2169. https://doi.org/10.1021/acssynbio.8b00145

Duellman, T., Doll, A., Chen, X., Wakamiya, R., Yang, J., 2017. dCas9-mediated transcriptional activation of tissue inhibitor of metalloproteinases. Met. Med. 4, 63–73. https://doi.org/10.2147/MNM.S146752

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., Church, G.M., 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat. Methods 10, 1116–1121. https://doi.org/10.1038/nmeth.2681

Fisher, A.L., Ohsako, S., Caudy, M., 1996. The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain. Mol. Cell. Biol. 16, 2670–2677. https://doi.org/10.1128/mcb.16.6.2670

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., Charpentier, E., 2016. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature 532, 517–521. https://doi.org/10.1038/nature17945

Gao, L., Cox, D.B.T., Yan, W.X., Manteiga, J.C., Schneider, M.W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., Zhang, F., 2017. Engineered Cpf1 variants with altered PAM specificities. Nat. Biotechnol. 35, 789–792. https://doi.org/10.1038/nbt.3900

Gao, X., Tsang, J.C.H., Gaba, F., Wu, D., Lu, L., Liu, P., 2014. Comparison of TALE designer transcription factors and the CRISPR/dCas9 in regulation of gene expression by targeting enhancers. Nucleic Acids Res. 42, e155. https://doi.org/10.1093/nar/gku836

Gao, X., Yang, J., Tsang, J.C.H., Ooi, J., Wu, D., Liu, P., 2013. Reprogramming to Pluripotency Using Designer TALE Transcription Factors Targeting Enhancers. Stem Cell Rep. 1, 183–197. https://doi.org/10.1016/j.stemcr.2013.06.002

Gao, Y., Xiong, X., Wong, S., Charles, E.J., Lim, W.A., Qi, L.S., 2016. Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. Nat. Methods 13, 1043–1049. https://doi.org/10.1038/nmeth.4042

Gao, Z., Harwig, A., Berkhout, B., Herrera-Carrillo, E., 2017. Mutation of nucleotides around the +1 position of type 3 polymerase III promoters: The effect on transcriptional activity and start site usage. Transcription 8, 275. https://doi.org/10.1080/21541264.2017.1322170

Garg, A., Lohmueller, J.J., Silver, P.A., Armel, T.Z., 2012. Engineering synthetic TAL effectors with orthogonal target sites. Nucleic Acids Res. 40, 7584–7595. https://doi.org/10.1093/nar/gks404

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., Qi, L.S., Kampmann, M.,

Weissman, J.S., 2014. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell 159, 647–661. https://doi.org/10.1016/j.cell.2014.09.029

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S., Qi, L.S., 2013. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell 154, 442–451. https://doi.org/10.1016/j.cell.2013.06.044

Goll, M.G., Bestor, T.H., 2005. Eukaryotic Cytosine Methyltransferases. Annu. Rev. Biochem. 74, 481–514. https://doi.org/10.1146/annurev.biochem.74.010904.153721

Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Dénervaud, N., Bucher, P., Trono, D., 2010. KRAB-Zinc Finger Proteins and KAP1 Can Mediate Long-Range Transcriptional Repression through Heterochromatin Spreading. PLoS Genet. 6, e1000869. https://doi.org/10.1371/journal.pgen.1000869

Halby, L., Champion, C., Sénamaud-Beaufort, C., Ajjan, S., Drujon, T., Rajavelu, A., Ceccaldi, A., Jurkowska, R., Lequin, O., Nelson, W.G., Guy, A., Jeltsch, A., Guianvarc'h, D., Ferroud, C., Arimondo, P.B., 2012. Rapid Synthesis of New DNMT Inhibitors Derivatives of Procainamide. ChemBioChem 13, 157–165. https://doi.org/10.1002/cbic.201100522

Hathaway, N.A., Bell, O., Hodges, C., Miller, E.L., Neel, D.S., Crabtree, G.R., 2012. Dynamics and memory of heterochromatin in living cells. Cell 149, 1447–1460. https://doi.org/10.1016/j.cell.2012.03.052

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., Gersbach, C.A., 2015. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. 33, 510–517. https://doi.org/10.1038/nbt.3199

Horvath, P., Romero, D.A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., Barrangou, R., 2008. Diversity, Activity, and Evolution of CRISPR Loci in Streptococcus thermophilus. J. Bacteriol. 190, 1401–1412. https://doi.org/10.1128/JB.01415-07

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A., 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J. Bacteriol. 169, 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science 337, 816–821. https://doi.org/10.1126/science.1225829

Kallimasioti-Pazi, E.M., Chathoth, K.T., Taylor, G.C., Meynert, A., Ballinger, T., Kelder, M.J.E., Lalevée, S., Sanli, I., Feil, R., Wood, A.J., 2018. Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. PLOS Biol. 16, e2005595. https://doi.org/10.1371/journal.pbio.2005595

Kay, S., Bonas, U., 2009. How Xanthomonas type III effectors manipulate the host plant. Curr. Opin. Microbiol. 12, 37–43. https://doi.org/10.1016/j.mib.2008.12.006

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., Maehr, R., 2015. Functional annotation of native enhancers with a Cas9–histone demethylase fusion. Nat. Methods 12, 401–403. https://doi.org/10.1038/nmeth.3325

Kim, H.K., Min, S., Song, M., Jung, S., Choi, J.W., Kim, Y., Lee, S., Yoon, S., Kim, H. (Henry), 2018. Deep learning improves prediction of CRISPR–Cpf1 guide RNA activity. Nat. Biotechnol. 36, 239–241. https://doi.org/10.1038/nbt.4061

Kim, H.K., Song, M., Lee, J., Menon, A.V., Jung, S., Kang, Y.-M., Choi, J.W., Woo, E., Koh, H.C., Nam, J.-W., Kim, H., 2017. *In vivo* high-throughput profiling of CRISPR–Cpf1 activity. Nat. Methods 14, 153–159. https://doi.org/10.1038/nmeth.4104

Kleinjan, D.A., Wardrope, C., Sou, S.N., Rosser, S.J., 2017. Drug-tunable multidimensional synthetic gene control using inducible degron-tagged dCas9 effectors. Nat. Commun. 8, 1–9. https://doi.org/10.1038/s41467-017-01222-y

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., Nureki, O., Zhang, F., 2015. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583–588. https://doi.org/10.1038/nature14136

Lee, J.G., Ha, C.H., Yoon, B., Cheong, S.-A., Kim, G., Lee, D.J., Woo, D.-C., Kim, Y.-H., Nam, S.-Y., Lee, S., Sung, Y.H., Baek, I.-J., 2019. Knockout rat models mimicking human atherosclerosis created by Cpf1-mediated gene targeting. Sci. Rep. 9, 1–9. https://doi.org/10.1038/s41598-019-38732-2

Lee, J.-H., Skalnik, D.G., 2008. Wdr82 Is a C-Terminal Domain-Binding Protein That Recruits the Setd1A Histone H3-Lys4 Methyltransferase Complex to Transcription Start Sites of Transcribed Human Genes. Mol. Cell. Biol. 28, 609–618. https://doi.org/10.1128/MCB.01356-07

Lee, J.-H., Skalnik, D.G., 2005. CpG-binding Protein (CXXC Finger Protein 1) Is a Component of the Mammalian Set1 Histone H3-Lys4 Methyltransferase Complex, the Analogue of the Yeast Set1/COMPASS Complex. J. Biol. Chem. 280, 41725–41731. https://doi.org/10.1074/jbc.M508312200

Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., Beisel, C.L., 2016. Identifying and Visualizing Functional PAM Diversity across CRISPR-Cas Systems. Mol. Cell 62, 137–147. https://doi.org/10.1016/j.molcel.2016.02.031

Liu, Q., Segal, D.J., Ghiara, J.B., Barbas, C.F., 1997. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. Proc. Natl. Acad. Sci. U. S. A. 94, 5525–5530. https://doi.org/10.1073/pnas.94.11.5525

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., Jaenisch, R., 2016. Editing DNA Methylation in the Mammalian Genome. Cell 167, 233-247.e17. https://doi.org/10.1016/j.cell.2016.08.056

Liu, X.S., Wu, H., Krzisch, M., Wu, X., Graef, J., Muffat, J., Hnisz, D., Li, C.H., Yuan, B., Xu, C., Li, Y., Vershkov, D., Cacace, A., Young, R.A., Jaenisch, R., 2018. Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. Cell 172, 979-992.e6. https://doi.org/10.1016/j.cell.2018.01.012

Liu, Y., Han, J., Chen, Z., Wu, H., Dong, H., Nie, G., 2017. Engineering cell signaling using tunable CRISPR–Cpf1-based transcription factors. Nat. Commun. 8, 1–8. https://doi.org/10.1038/s41467-017-02265-x

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M., Misteli, T., 2010. Regulation of Alternative Splicing by Histone Modifications. Science 327, 996–1000. https://doi.org/10.1126/science.1184208

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., Joung, J.K., 2013a. CRISPR RNA–guided activation of endogenous human genes. Nat. Methods 10, 977–979. https://doi.org/10.1038/nmeth.2598

Maeder, M.L., Linder, S.J., Reyon, D., Angstman, J.F., Fu, Y., Sander, J.D., Joung, J.K., 2013b. Robust, synergistic regulation of human gene expression using TALE activators. Nat. Methods 10, 243–245. https://doi.org/10.1038/nmeth.2366

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., Church, G.M., 2013. RNA-Guided Human Genome Engineering via Cas9. Science 339, 823–826. https://doi.org/10.1126/science.1232033

Margolin, J.F., Friedman, J.R., Meyer, W.K.-H., Vissing, H., Thiesen, H.-J., Rauscher, F.J., 1994. Kruppel-Associated Boxes are Potent Transcriptional Repression Domains. Proc. Natl. Acad. Sci. U. S. A. 91, 4509–4513.

Matharu, N., Rattanasopha, S., Tamura, S., Maliskova, L., Wang, Y., Bernard, A., Hardin, A., Eckalbar, W.L., Vaisse, C., Ahituv, N., 2019. CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. Science 363. https://doi.org/10.1126/science.aau0629

Metzger, E., Wissmann, M., Yin, N., Müller, J.M., Schneider, R., Peters, A.H.F.M., Günther, T., Buettner, R., Schüle, R., 2005. LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. Nature 437, 436–439. https://doi.org/10.1038/nature04020

Mitchell, L.A., Cai, Y., Taylor, M., Noronha, A.M., Chuang, J., Dai, L., Boeke, J.D., 2013. Multichange Isothermal Mutagenesis: A New Strategy for Multiple Site-Directed Mutations in Plasmid DNA. ACS Synth. Biol. 2, 473–477. https://doi.org/10.1021/sb300131w

Morbitzer, R., Römer, P., Boch, J., Lahaye, T., 2010. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. Proc. Natl. Acad. Sci. 107, 21617–21622. https://doi.org/10.1073/pnas.1013133107

Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P., Lu, T.K., 2014. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. Mol. Cell 54, 698–710. https://doi.org/10.1016/j.molcel.2014.04.022

Oakes, B.L., Nadler, D.C., Flamholz, A., Fellmann, C., Staahl, B.T., Doudna, J.A., Savage, D.F., 2016. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. Nat. Biotechnol. 34, 646–651. https://doi.org/10.1038/nbt.3528

Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., Nakatani, Y., 1996. The Transcriptional Coactivators p300 and CBP Are Histone Acetyltransferases. Cell 87, 953–959. https://doi.org/10.1016/S0092-8674(00)82001-2

Orioli, A., Pascali, C., Quartararo, J., Diebel, K.W., Praz, V., Romascano, D., Percudani, R., van Dyk, L.F., Hernandez, N., Teichmann, M., Dieci, G., 2011. Widespread occurrence of non-canonical transcription termination by human RNA polymerase III. Nucleic Acids Res. 39, 5499–5512. https://doi.org/10.1093/nar/gkr074

Park, H.M., Liu, H., Wu, J., Chong, A., Mackley, V., Fellmann, C., Rao, A., Jiang, F., Chu, H., Murthy, N., Lee, K., 2018. Extension of the crRNA enhances Cpf1 gene editing in vitro and in vivo. Nat. Commun. 9, 1–12. https://doi.org/10.1038/s41467-018-05641-3

Park, J., Bae, S., 2018. Cpf1-Database: web-based genome-wide guide RNA library design for gene knockout screens using CRISPR-Cpf1. Bioinformatics 34, 1077–1079. https://doi.org/10.1093/bioinformatics/btx695

Pavletich, N.P., Pabo, C.O., 1991. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science 252, 809–817. https://doi.org/10.1126/science.2028256

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., Guilak, F., Crawford, G.E., Reddy, T.E., Gersbach, C.A., 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat. Methods 10, 973–976. https://doi.org/10.1038/nmeth.2600

Perumbeti, A., Higashimoto, T., Urbinati, F., Franco, R., Meiselman, H.J., Witte, D., Malik, P., 2009. A novel human gamma-globin gene vector for genetic correction of sickle cell anemia in a humanized sickle mouse model: critical determinants for successful correction. Blood 114, 1174–1185. https://doi.org/10.1182/blood-2009-01-201863

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A., 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control

of gene expression. Cell 152, 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022

Rivera, V.M., Berk, L., Clackson, T., 2012. Dimerizer-mediated regulation of gene expression in vivo. Cold Spring Harb. Protoc. 2012, 821–824. https://doi.org/10.1101/pdb.prot070144

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C., Pfenning, A.R., Wang, X., Claussnitzer, M., Yaping Liu, Coarfa, C., Alan Harris, R., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., David Hawkins, R., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Scott Hansen, R., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., Jager, P.L.D., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330. https://doi.org/10.1038/nature14248

Ryan, R.F., Schultz, D.C., Ayyanathan, K., Singh, P.B., Friedman, J.R., Fredericks, W.J., Rauscher, F.J., 1999. KAP-1 Corepressor Protein Interacts and Colocalizes with Heterochromatic and Euchromatic HP1 Proteins: a Potential Role for Krüppel-

Associated Box–Zinc Finger Proteins in Heterochromatin-Mediated Gene Silencing. Mol. Cell. Biol. 19, 4366–4378. https://doi.org/10.1128/MCB.19.6.4366

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., Siksnys, V., 2011. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res. 39, 9275–9282. https://doi.org/10.1093/nar/gkr606

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., Rauscher, F.J., 2002. SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. 16, 919–932. https://doi.org/10.1101/gad.973302

Shi, Yujiang, Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., Shi, Yang, 2004. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. Cell 119, 941–953. https://doi.org/10.1016/j.cell.2004.12.012

Snowden, A.W., Gregory, P.D., Case, C.C., Pabo, C.O., 2002. Gene-Specific Targeting of H3K9 Methylation Is Sufficient for Initiating Repression In Vivo. Curr. Biol. 12, 2159–2166. https://doi.org/10.1016/S0960-9822(02)01391-X

Sripathy, S.P., Stevens, J., Schultz, D.C., 2006. The KAP1 Corepressor Functions To Coordinate the Assembly of De Novo HP1-Demarcated Microenvironments of Heterochromatin Required for KRAB Zinc Finger Protein-Mediated Transcriptional Repression. Mol. Cell. Biol. 26, 8623–8638. https://doi.org/10.1128/MCB.00487-06

Stevens, A.J., Brown, Z.Z., Shah, N.H., Sekar, G., Cowburn, D., Muir, T.W., 2016. Design of a Split Intein with Exceptional Protein Splicing Activity. J. Am. Chem. Soc. 138, 2162–2165. https://doi.org/10.1021/jacs.5b13528

Streubel, J., Blücher, C., Landgraf, A., Boch, J., 2012. TAL effector RVD specificities and efficiencies. Nat. Biotechnol. 30, 593–595. https://doi.org/10.1038/nbt.2304

Suzuki, M.M., Kerr, A.R.W., De Sousa, D., Bird, A., 2007. CpG methylation is targeted to transcription units in an invertebrate genome. Genome Res. 17, 625–631. https://doi.org/10.1101/gr.6163007

Tak, Y.E., Kleinstiver, B.P., Nuñez, J.K., Hsu, J.Y., Horng, J.E., Gong, J., Weissman, J.S., Joung, J.K., 2017. Inducible and multiplex gene regulation using CRISPR–Cpf1-based transcription factors. Nat. Methods 14, 1163–1166. https://doi.org/10.1038/nmeth.4483

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., Vale, R.D., 2014. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. Cell 159, 635–646. https://doi.org/10.1016/j.cell.2014.09.039

Teif, V.B., Rippe, K., 2010. Statistical–mechanical lattice models for protein–DNA binding in chromatin. J. Phys. Condens. Matter 22, 414105. https://doi.org/10.1088/0953-8984/22/41/414105

Thakore, P.I., D'Ippolito, A.M., Song, L., Safi, A., Shivakumar, N.K., Kabadi, A.M., Reddy, T.E., Crawford, G.E., Gersbach, C.A., 2015. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. Nat. Methods 12, 1143–1149. https://doi.org/10.1038/nmeth.3630

Tóth, E., Czene, B.C., Kulcsár, P.I., Krausz, S.L., Tálas, A., Nyeste, A., Varga, É., Huszár, K., Weinhardt, N., Ligeti, Z., Borsy, A.É., Fodor, E., Welker, E., 2018. Mb- and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants. Nucleic Acids Res. 46, 10272–10285. https://doi.org/10.1093/nar/gky815

Truong, V.A., Hsu, M.-N., Kieu Nguyen, N.T., Lin, M.-W., Shen, C.-C., Lin, C.-Y., Hu, Y.-C., 2019. CRISPRai for simultaneous gene activation and inhibition to promote stem cell chondrogenesis and calvarial bone regeneration. Nucleic Acids Res. 47, e74–e74. https://doi.org/10.1093/nar/gkz267

Tsukamoto, T., Sakai, E., Iizuka, S., Taracena-Gándara, M., Sakurai, F., Mizuguchi, H., 2018. Generation of the Adenovirus Vector-Mediated CRISPR/Cpf1 System and the Application for Primary Human Hepatocytes Prepared from Humanized Mice with Chimeric Liver. Biol. Pharm. Bull. 41, 1089–1095. https://doi.org/10.1248/bpb.b18-00222

Tu, M., Lin, L., Cheng, Y., He, X., Sun, H., Xie, H., Fu, J., Liu, C., Li, J., Chen, D., Xi, H., Xue, D., Liu, Q., Zhao, J., Gao, C., Song, Z., Qu, J., Gu, F., 2017. A 'new lease of life': FnCpf1 possesses DNA cleavage activity for genome editing in human cells. Nucleic Acids Res. 45, 11295–11304. https://doi.org/10.1093/nar/gkx783

Wang, J.-C., 2005. Finding primary targets of transcriptional regulators. Cell Cycle Georget. Tex 4, 356–358. https://doi.org/10.4161/cc.4.3.1521

Wright, A.V., Sternberg, S.H., Taylor, D.W., Staahl, B.T., Bardales, J.A., Kornfeld, J.E., Doudna, J.A., 2015. Rational design of a split-Cas9 enzyme complex. Proc. Natl. Acad. Sci. 112, 2984–2989. https://doi.org/10.1073/pnas.1501698112

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., Ishitani, R., Zhang, F., Nureki, O., 2016. Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell 165, 949–962. https://doi.org/10.1016/j.cell.2016.04.003

Yang, H., Wang, H., Shivalila, C.S., Cheng, A.W., Shi, L., Jaenisch, R., 2013. One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. Cell 154, 1370–1379. https://doi.org/10.1016/j.cell.2013.08.022

Yang, S., Zheng, X., Lu, C., Li, G.-M., Allis, C.D., Li, H., 2016. Molecular basis for oncohistone H3 recognition by SETD2 methyltransferase. Genes Dev. 30, 1611–1616. https://doi.org/10.1101/gad.284323.116

Yosef, I., Goren, M.G., Qimron, U., 2012. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res. 40, 5569–5576. https://doi.org/10.1093/nar/gks216

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., Koonin, E.V., Zhang, F., 2015a. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771. https://doi.org/10.1016/j.cell.2015.09.038

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., Koonin, E.V., Zhang, F., 2015b. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771. https://doi.org/10.1016/j.cell.2015.09.038

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., Wu, W.Y., Scott, D.A., Severinov, K., van der Oost, J., Zhang, F., 2017. Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. Nat. Biotechnol. 35, 31–34. https://doi.org/10.1038/nbt.3737

Zetsche, B., Volz, S.E., Zhang, F., 2015c. A split-Cas9 architecture for inducible genome editing and transcription modulation. Nat. Biotechnol. 33, 139–142. https://doi.org/10.1038/nbt.3149

Zhang, D., Zhang, H., Li, T., Chen, K., Qiu, J.-L., Gao, C., 2017. Perfectly matched 20-nucleotide guide RNA sequences enable robust genome editing using high-fidelity SpCas9 nucleases. Genome Biol. 18, 191. https://doi.org/10.1186/s13059-017-1325-9

Zhang, X., Wang, Jingman, Cheng, Q., Zheng, X., Zhao, G., Wang, Jin, 2017. Multiplex gene regulation by CRISPR-ddCpf1. Cell Discov. 3, 17018. https://doi.org/10.1038/celldisc.2017.18

Zhang, Xin, Wang, W., Shan, L., Han, L., Ma, S., Zhang, Y., Hao, B., Lin, Y., Rong, Z., 2018. Gene activation in human cells using CRISPR/Cpf1-p300 and CRISPR/Cpf1-SunTag systems. Protein Cell 9, 380–383. https://doi.org/10.1007/s13238-017-0491-6

Zhang, Xuhua, Xu, L., Fan, R., Gao, Q., Song, Yunfeng, Lyu, X., Ren, J., Song, Yongping, 2018. Genetic editing and interrogation with Cpf1 and caged truncated pre-tRNA-like crRNA in mammalian cells. Cell Discov. 4, 1–10. https://doi.org/10.1038/s41421-018-0035-0

Zhang, Y., Long, C., Li, H., McAnally, J.R., Baskin, K.K., Shelton, J.M., Bassel-Duby, R., Olson, E.N., 2017. CRISPR-Cpf1 correction of muscular dystrophy mutations in human cardiomyocytes and mice. Sci. Adv. 3. https://doi.org/10.1126/sciadv.1602814